AD_____

AWARD NUMBER: W81XWH-04-1-0570

TITLE: Endocrine Therapy of Breast Cancer

PRINCIPAL INVESTIGATOR: Robert Clarke, Ph.D.

CONTRACTING ORGANIZATION: Georgetown University
Washington, DC 20007

REPORT DATE: June 2009

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
1 June 2009	Final	17 May 2004 – 16 May 2009

4. TITLE AND SUBTITLE

Endocrine Therapy of Breast Cancer

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-04-1-0570

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Robert Clarke, Ph.D.

E-Mail: clarker@georgetown.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Georgetown University
Washington, DC 20007

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

A controversy in the treatment of estrogen receptor positive (ER+) breast cancers is whether an aromatase inhibitor, e.g., letrozole (LET) or TAM should be given as first line endocrine therapy. Unfortunately, response rates are lower, and response durations are shorter, on crossover than when these agents are given as first line therapies, e.g., ~40% of tumors show cross-resistance to TAM or an aromatase inhibitor on crossover. Only 50% of ER+ tumors respond to endocrine therapy. Currently, we fail to predict endocrine responsiveness in about 66% of ER+/PgR- (progesterone receptor), 55% of ER-/PgR+, and 25% of ER+/PgR+ tumors. In this Clinical Translational Research Award, we hypothesize that our analytical methods (including new methods we will develop/test) can extract expression profiles of breast tumors that define their responsiveness (sensitive vs. resistant) to endocrine therapies. These profiles will support neural network and biostatistical classifiers or committee machines that predict each tumor's endocrine responsiveness. Our objectives are to array breast cancer cases, build classifiers of endocrine responsiveness (using microarray data), and validate these classifiers in independent data sets. Genes will be further studied using cellular and molecular methods, and their role as therapeutic targets explored. In the long term, we will design custom arrays for use in clinical practice.

15. SUBJECT TERMS
Antiestrogen, aromatase inhibitor, anastrazole, bioinformatics, biomarkers, biostatistics, breast cancer, class prediction, clinical trial, computer science, engineering, immunohistochemistry, letrozole, microarrays, molecular profiling, neural networks, recurrence, resistance, tamoxifen

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
U	U	U	UU	90	19b. TELEPHONE NUMBER *(include area code)*

TABLE OF CONTENTS

1. Wang, et al., *Br J Cancer,* 98: 1023-1028, 2008.
2. Shajahan, et al., Chapter 8. In: *Breast Cancer: Prognosis, Treatment, and Prevention*, pp137-156, 2008.
3. Chen, et al., *BMC Bioinformatics*, 9:416 (16 pages as published on-line), 2008
4. Riggins, et al., *Cancer Res,* 68: 8908-8917, 2008.
5. Zhang, et al., *Bioinformatics*, 25: 526-532, 2009.
6. Clarke, et al., *J Steroid Biochem Mol Biol*, 114: 8-20, 2009.
7. Xuan, et al., *Front Biosci*, in press, 2009.
8. Cavalli, et al., *Breast Cancer Res Treat*, in press, 2009.

INTRODUCTION

Endocrine therapy is often the least toxic and most effective treatment for hormone receptor positive invasive breast cancer. Such therapy includes antiestrogens (tamoxifen, fulvestrant) and aromatase inhibitors (AI; *e.g.*, anastrozole, letrozole, exemestane). Tamoxifen (TAM) increases disease free and overall survival in the adjuvant setting, reduces the incidence of estrogen receptor positive disease (ER+; unless otherwise noted ER=ERα) in high-risk women, and reduces the rate of bone loss secondary to osteoporosis in postmenopausal women [1,2]. Aromatase inhibitors are effective only in the absence of functioning ovaries - TAM can be used regardless of menopausal status. Recent studies suggest that AIs may be superior to TAM in the adjuvant treatment of postmenopausal women with ER+ breast cancer; other studies report higher overall response rates with AI *vs.* TAM as first line therapy in the metastatic setting. Thus, a controversy in the management of patients with ER+ disease is whether an aromatase inhibitor or TAM should be given as first line endocrine therapy [3-9].

In this Clinical Translational Research award, we propose to build classifiers that accurately separate antiestrogen responsive from antiestrogen resistant/unresponsive breast tumors and begin to assist in the direction of specific endocrine treatments (antiestrogen *vs.* aromatase inhibitor) to *individual* patients. We hypothesize that endocrine responsiveness is affected by a gene network, rather than the activity of only one or two genes or signaling pathways [10-12]. Since the key components of such a network are unknown, we must study 10,000s of genes. We will use Affymetrix GeneChips. We will not identify mutational events, the presence of mRNA splice variants, or post-translational protein modifications. However, these factors have major effects on the transcriptome and their "footprints" should be identified by expression microarrays.

BODY

Overview: We will build classifiers with the ultimate goal of separating antiestrogen sensitive from antiestrogen resistant breast tumors and begin to assist in the direction of specific endocrine treatments (antiestrogen *vs.* aromatase inhibitor) to *individual* patients. To achieve this goal, and consistent with a CTR award, we will conduct a 4-year, prospective, neoadjuvant study with an aromatase inhibitor (AI) or antiestrogen (Tamoxifen; TAM) as the *only* systemic therapy. We will obtain molecular profiles from Affymetrix GeneChips and further develop and apply our innovative bioinformatic and biostatistic methods to explore these high dimensional data sets and build/validate new classifiers. A more accurate predictor of endocrine responsiveness would have widespread clinical use, allowing women and physicians to make more individualized and appropriate treatment decisions. For example, patients with tumors predicted to be resistant to antiestrogens and/or aromatase inhibitors would be strong candidates for an early intervention with cytotoxic chemotherapy.

In most predictive/prognostic marker studies investigators focus on a *single* factor and whether they obtain a p-value that reaches conventional statistical significance. Our approach is different because we will determine whether we can also find joint gene subsets that can separate patients into sufficiently distinct groups that should differ in their treatment. We will **(1)** analyze >33,000 genes on retrospective and prospective material, **(2)** apply (and develop/improve) new biostatistical and bioinformatic methods to identify potentially informative "biomarkers," **(3)** build neural network and biostatistical model classifiers, **(4)** evaluate the joint discriminant power of selected genes concurrently rather than as single biomarkers, **(5)** focus on prediction for individual patients where the assessment of a p-value is less important than the classification rate of our predictors, **(6)** validate the classifiers in independent data sets, and **(7)** explore the ability of predictors to refine the targeting of *specific* endocrine therapies.

Evidence has begun to accumulate suggesting that an AI might be a more effective first line endocrine therapy for some breast cancer patients than TAM. These data have generated considerable interest and controversy, in part

because unlike TAM, there are no long term studies with AIs where definitive survival data are available from up to 20 years of clinical follow-up. Our study could provide new and innovative insights into how to approach the more effective targeting of specific endocrine therapies to individual patients.

Specific Aims

We will complete two clinical studies and collect gene expression profiles from which to build predictors of endocrine responsiveness. Predictors will be built in Specific Aim 2 and validated in Specific Aim 3.

AIM 1: Clinical Studies - **Clinical Study-1** (retrospective) is of pretreatment, single, frozen samples where we will compare the molecular profiles of tumors that recurred on TAM with those of tumors that did not recur. **Clinical Study-2** is a prospective study of breast tumor samples from patients treated with neoadjuvant TAM or AI.

AIM 2: We will develop and apply novel bioinformatics and biostatistics to discover gene subsets that define the molecular differences between endocrine sensitive and resistant breast tumors.

AIM 3: We will test, optimize, and validate the performance of the classifiers from Aim 2 in retrospective studies of human breast tumors.

KEY RESEARCH ACCOMPLISHMENTS

As noted in previous reports, progress on the clinical goals for this award was greatly delayed because of the time taken to obtain DOD approval of our preexisting institutionally approved IRBs at Georgetown University and at the University of Edinburgh. All institutionally approved protocols and requested material were submitted to the DOD in July 2004; additional information was requested by the DOD several months later and submitted in November 2004. We did not receive final approval to proceed with the clinical studies until March 2005. Much of this delay seems to have been entirely unavoidable (see prior reports). While this affected progress on some tasks, we made significant strides and advanced the research program in each of the three primary tasks. Our development of new analytical procedures and the research infrastructure has been largely completed, although we will continue to develop these as they have applicability beyond the work in this study and because the technologies in the field continue to evolve. We will also continue (beyond the ending of this award) to follow patient performance and to update our clinical records for the prospective study and, as appropriate, reanalyze and report further progress. Publications supported since the commencement of this no-cost extension are listed under "Reportable Outcomes"; these constitute some of our major accomplishments in the past year. Since we are now working on publishing the data from these two clinical studies (retrospective; prospective), these and other key research accomplishments are presented in more detail below.

Progress on our Statement of Work (narrative)

- **TASK 1.** Array breast tumor samples from Clinical Studies 1 (retrospective) and 2 (prospective)

We have obtained breast specimens from breast cancer patients treated with endocrine therapy (or not, *i.e.*, surgery and radiation only in selected retrospective cases) as described in the original application. These specimens represent a mix of the initial prospective and retrospective specimens. All of these specimens have been fully analyzed and annotated by the study pathologist and total RNA has been extracted, and either stored, labeled and/or arrayed. Thus, we have also completed the assessment of tissue, RNA, and microarray data quality control as appropriate.

We requested that the specimens be sent independent of the clinical information, so that we could adequately and appropriately randomize the RNA preparation, labeling, and hybridization and minimize any operator-induced or technology-induced bias. All specimens were processed using our standard operating procedures; each manipulation being performed by the same individual to further reduced inter-operator variability. Details of the methods, quality control measures, and general experimental approaches have been described in detail in earlier annual reports.

We have found the data from the two clinical studies in this CTR to be particularly useful in supporting other studies that are ongoing in the laboratory. For example, these data have been used to support R01 applications on genes we identified and described in the preliminary data for this application. One of these, which is focused on components of a gene network associated with endocrine resistance in breast cancer, was supported, in part, by preliminary data obtained by analysis of the data generated specifically in this CTR application (R01-CA131465; awarded in June 2009). Indeed, we have found the unique datasets from this CTR to be particularly powerful for exploring the potential of individual genes to act as biomarkers of endocrine resistance (in addition to the neural network classifiers described later in this report). These genes can be obtained from several sources including as individual genes selected by analysis of the data in the retrospective and/or prospective clinical studies, and/or from other work in our laboratory.

As in prior years, we have also used these data to provide preliminary data on gene expression values that have led to our colleagues initiating other studies directed at developing therapeutic strategies to target individual genes we have identified from within this data set or from other sources. One of the genes described in our initial preliminary data and studied further as an individual gene in the data from this CTR has led directly to the very recent identification of a potential new breast cancer drug. This biomarker gene (XBP1) was discussed in the original application and in subsequent annual reports, and we have now identified a small molecule inhibitor of its activation that seems to have activity against breast cancer cells. While this molecule was developed primarily using institutional funds (and so it is not discussed further in this report), it represents a definitive step towards addressing one of the longer term goals of this award.

While, as noted in prior reports, we had prioritized the analysis of data from the retrospective study material because these have definitive clinical outcomes (survival), as noted for Tasks 2/3 (below), we have also analyzed the material from our prospective study. We will continue to collect information on recurrence status of the patients accrued to the prospective study because we expect informative events that will strengthen our study to continue to occur. Since many endocrine treated breast cancers tend to recur in later years, it is to be expected that we may obtain our most valuable data (recurrence) after this award has ended. While we report our initial analyses of the prospective study (see Tasks 2/3), we expect to conduct further analysis in the future. We are fully committed to generating the most complete analysis we can perform, and we believe that we have made very strong progress in this task. We will continue to report the outcomes of future research on these data sets in the literature, and to submit these publications to the DOD. We will bear the costs of these future analyses and clinical follow up studies from other sources as appropriate.

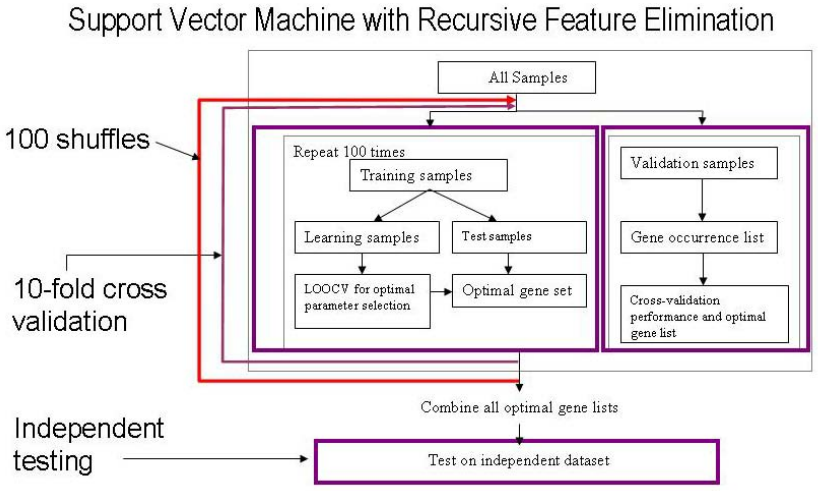
We have completed almost all of the work on this Task as anticipated in prior reports. We have now also completed a major analysis of the retrospective study (Clinical Study-1) described in the approved report submitted last year (2008) - see Tasks 2/3 below. For the prospective study (Clinical Study-2), we will continue to obtain outcomes data and to update the database and research analyses as the clinical information dictates. Analyses are also provided below (Tasks 2/3).

• **TASK 2.** Store, process, and train/optimize classifiers from gene expression microarray data (already modified to reflect our adoption of caArray and other caBIG tools in prior reports)

We have made significant progress on addressing the infrastructure goals in this task, largely as a consequence of our involvement in the National Cancer Institute Center caBIG project (as also noted in prior reports). The PI (Dr. Clarke) also leads the Lombardi Comprehensive Cancer Center's (LCCC) caBIG team and we have been actively involved in the development of caArray (caBIG's grid-enabled, MIAME compliant, microarray database). Our work in the development of the informatics infrastructure that currently houses the clinical and microarray data generated in this CTR, and the analytical tools we have developed in part with support from this CTR (as proposed in the original application) have been particularly productive. Most notable, our work in this area has led to the successful award of a caBIG *In Silico* Center of Excellence (RFP CA-S09-094; July 2009-2012; Dr. Clarke is the PI/Scientific Director; Dr. Madhavan is the other co-PI and the Project Manager): a subcontract to our colleagues at Virginia Tech is also included (same laboratories as were supported in this CTR). This new award will allow us to continue to advance and improve the clinical-translational research infrastructure development that we described in our prior reports. The research infrastructure will be further developed to support cancer research within the LCCC and shared fully with the broader research community. Our adoption of caBIG and our activities in this large federally funded research community also allows us to share our expertise and tools with many other cancer researchers.

We also continue to further develop and optimize our data analysis algorithms, with particular success in the design of new approaches for network analysis. We have found approaching this goal to be realistic in a much shorter time-frame than initially expected and have now published several relevant articles. We also continue to improve our existing algorithms for data analysis. Relevant publications in this area are also included below in the section "Reportable Outcomes."

We have now acquired the data for analysis of the endocrine therapies and outcomes. We have gone substantially beyond the initial analysis reported last year. We now report optimized and validated classifiers of TAM responsiveness (Clinical Study-1). Please note that we include initial validation studies (Task 3) here also for the sake of brevity and clarity. We also provide analysis of the data from the prospective study (Clinical Study-2). The data analysis format follows that reported last year; it is shown again here to allow reviewers to follow the procedures.



The data analysis design is shown above (LOOCV = leave-one-out cross validation).

From this scheme, testing and optimization of the classifier is performed using 10-fold crossvalidation within the BC030280 data set for the analysis of Clinical Study-1.

We have determined that the existing published data sets we used previously are somewhat useful (some of these concerns were indicated in last year's report) but we could not use at least two of these to validate our endocrine responsiveness classifiers (Task 3). The only data set available for such validation has been published by Loi *et al.* [13]. Unfortunately, the dataset they used was created by combining data from multiple centers/studies. To address our concerns with the Loi dataset and to provide a more rigorous validation dataset, we selected a subset of their data, where the patients, their treatments, outcomes, and follow-up more closely match those in the BC030280 data set. Thus, we now have our single institution dataset (BC030280 data set from the retrospective clinical study is probably still the only such dataset with meaningful follow-up), and an adequate validation data set (validation is a key aspect of Task 3).

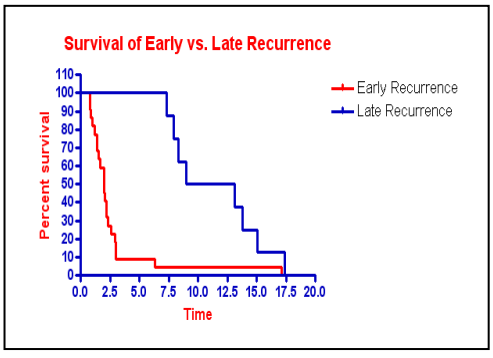
In our analyses, the datasets now enable us to ask, for the first time, important and previously unanswered questions in clinical endocrine resistance. In this report, we describe, for the first time, an innovative analysis in which we separate early recurrent disease from late recurrent disease (other analyses will continue as we test other hypotheses in these data sets). The hypothesis being tested is that early recurrences are likely to be simply reflective of poor prognosis, whereas later recurrences have a different biology that is more reflective of true endocrine resistance. If we cannot accurately separate these two groups, then the underlying biology of early vs. late recurrent disease is probably the same. However, either outcome is almost equally important and informative, and of great significance both for the direction of therapy and understanding for understanding the biology of endocrine sensitive and resistant breast cancer. For example, patients that will recur early are unlikely to receive major benefit from TAM therapy. Those at risk of recurring later, may require more intensive follow up for a prolonged period. Understanding the biology driving the late recurrences could also lead to the development of new drugs to prevent or delay further the emergence of recurrent disease.

For Clinical Study-1, the predictive classifiers were built, trained/optimized using the BC030280 data set.

For each analysis, we set challenging recurrence endpoints: ≤3 yrs *vs.* late ≥15 yrs (we studied only *distant* recurrent breast cancers). Performance was evaluated against a broad series of preset benchmarks (expanded substantially from those described in last year's report), *i.e.*, ≥70% performance in each of accuracy, sensitivity, specificity (from the receiver operating characteristic curve; ROC); AUC≥0.8; ≥0.70 negative predictive value (NPV); ≥0.70 positive predictive value (PPV). We also estimated the hazard ratios and visualized the data using standard Kaplan-Meier plots and required models to achieve log rank test $p<0.05$ and HR≥2.0.

Results for the optimized classifier BC030280 of early (≤3 yrs) *vs.* late (≥15 yrs) recurrence on TAM.

Accuracy	Specificity	Sensitivity	AUC	PPV	NPV	P-value	Hazard Ratio
0.93	1.00	0.80	0.82	1.00	0.91	0.0006	3.28

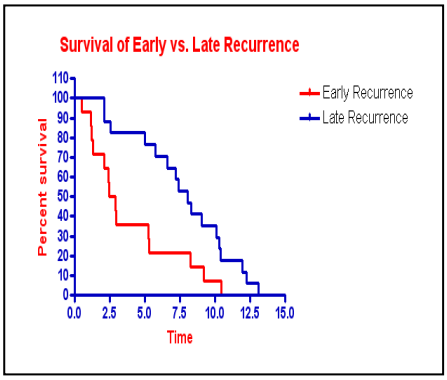


As evident from the data, the optimized classifier exceeded or met all requirements.

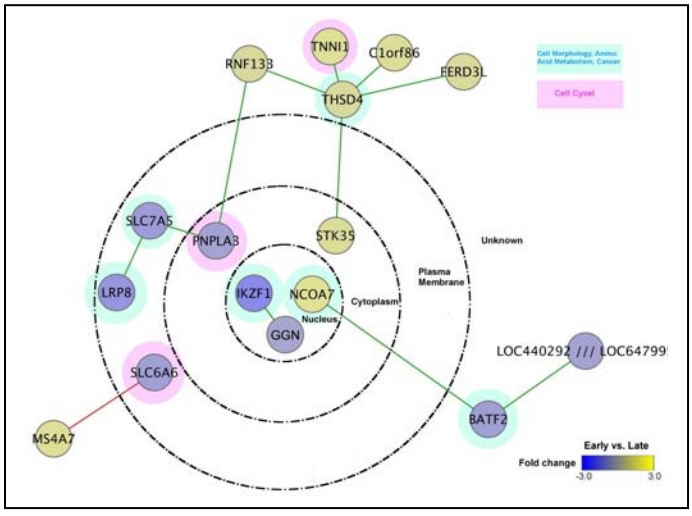
As we have previously noted, independent validation is essential to account for over-fitting and to assess the likely robustness of a classifier across other data sets [14]. Thus, independent validation was done in a subset of the published dataset of Loi, S., Haibe-Kains, B., Desmedt, C., Wirapati, P., Lallemand, F., Tutt, A. M., Gillet, C., Ellis, P., Ryder, K., Reid, J. F., Daidone, M. G., Pierotti, M. A., Berns, E. M., Jansen, M. P., Foekens, J. A., Delorenzi, M., Bontempi, G., Piccart, M. J. and Sotiriou, C. *Predicting prognosis using molecular profiling in estrogen receptor-positive breast cancer treated with tamoxifen.* BMC Genomics, *9:* 239, 2008

Results for validation of the optimized classifier BC030280 of early (≤3 yrs) *vs.* late (≥15 yrs) recurrence on TAM (Loi data set).

Accuracy	Specificity	Sensitivity	AUC	PPV	NPV	P-value	Hazard Ratio
0.74	0.75	0.74	0.83	0.82	0.64	0.01	2.35

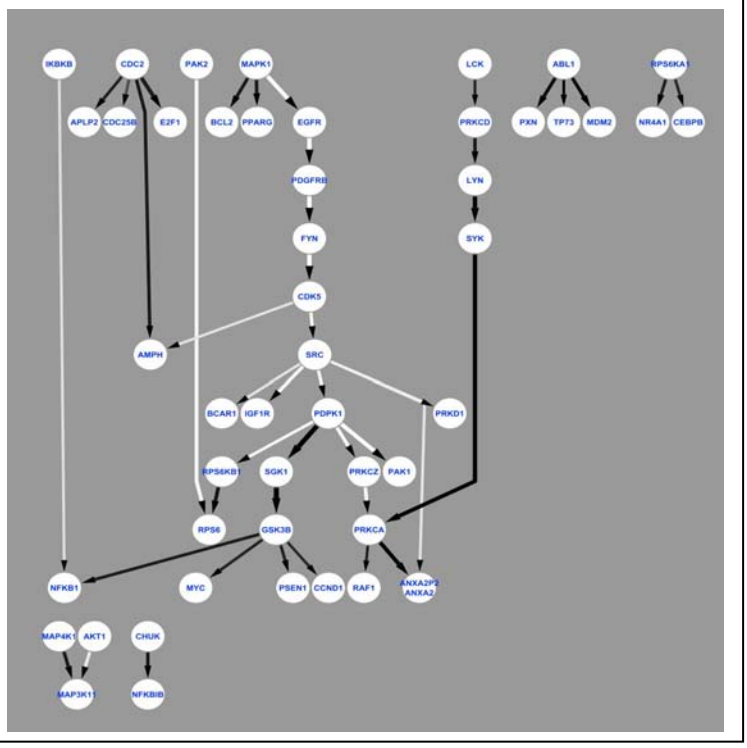


As evident from the data, the optimized classifier exceeded all but one of the stringent performance requirements. These data provide definitive validation of the performance of the classifier.



The data from the studies described above provide compelling evidence that there are fundamental differences in the biology of ER+ breast cancer that will recur early (≤3 yrs) vs. those that will recur much later (≥15 yrs). We have begun to explore these data to understand the biological differences that may explain why these two groups of recurrent breast cancers are different. To date, we have applied our Differential Dependency Network analysis to extract small preliminary subnetworks. While these studies are still in progress, the figure shown here reports our first model. We have no immediate interpretation of these genes from a mechanistic study, as this initial model implicates many genes not previously known to be associated with breast cancer. However, altered function of nuclear receptors is strongly implicated by the inclusion of NCOA7 (nuclear receptor coactivator 7; ERAP140), which binds ER-alpha and mediates gene regulation [15]. IKZF1 has very recently been associated with poor outcome in acute lymphoblastic leukemia [16].

To understand the mechanistic differences between breast tumors that recur on TAM (at any time point) and those that do not, we have begun to explore the differential expression of genes across the entire BC030280 dataset. Focusing initially upon genes known to encode kinases, and applying a novel method (currently undergoing further development prior to publication), we have already obtained a preliminary insight into signaling within these breast tumors. The fig to the right shows the outcome from this first-pass analysis. Many known kinases and their targets are linked throughout this pathway. Additional studies will continue as we attempt to understand mechanism and to identify new targets for drug discovery.



We have fewer specimens and less definitive outcomes data for the prospective AI study incorporated into Clinical Study-2, and there are no adequate published data sets for validation. Nonetheless, we have attempted an initial classifier trained by crossvalidation. While we would not expect to be able to meet the rather stringent panel of performance measures used for the study above on Clinical Study-1, we have explored prediction potential with the data currently available. For this analysis, we used only data from the AI treated cases. Since the data set is smaller than that for the TAM only retrospective study, cases were defined as “good” (partial response + minimal response) and “bad” (no change + progressive disease) based on clinical response.

Results from the initial model using all samples are deceptively good and almost certainly reflect an overfitting of the model to the data.

Accuracy	Sensitivity	Specificity
100%	100%	100%

We then used a leave-one-out crossvalidation to address possible overfitting. As is evident from the results, performance of an AI-only model is limited, only the measure of sensitivity approaches our performance benchmarks. This likely reflects the fewer samplers available for study. While we obtained the specimens we requested for this application, we are currently exploring the possibility of obtaining additional specimens/data from our collaborators in Edinburgh, *i.e.*, Drs. Dixon and Miller. Such data will likely dramatically improve performance of the AI classifier. We are also in the process of using the TAM alone predictor from Clinical Study-1 to predict outcomes from the TAM and AI specimens in Clinical Trial-2.

Accuracy	Sensitivity	Specificity
58%	67%	33%

Interestingly, there is no overlap in the listing of differentially expressed genes from the early vs. late TAM recurrence data sets and the AI responders vs. non-responders data. This observation suggests that those genes responsible for responsiveness to AIs are not the same as those directing TAM responsiveness. Overall, we believe that progress on Tasks 2/3 is fully consistent with our initial goals.

• **TASK 3.** Test, optimize, and validate the performance of the classifiers from Aim 2 in retrospective studies of human breast tumors.

We proposed to rank and prioritize selected joint genes from RNA classifier built and optimized in Task 2 (above) and retrain/reoptimize the initial neural network classifier. Finally, we will validate IHC classifier on independent data sets (data sets not used to build and train the classifiers).

Please note that, for the sake of clarity, the outcomes from independent validation (in the reconstructed Loi data set) are incorporated above in the report for Task 2.

As acknowledged in prior reports, we remain unable to move this later part of Task 3 (IHC studies) substantially forward on the timeframe as initially proposed because of the delays in getting approval to work with the clinical specimens (this aspect of Task 3 cannot begin until Tasks 1 and 2 are almost complete). Thus, we will address the IHC studies in the future at our expense as necessary and appropriate. Nonetheless, we have been able to use published data to validate our classifiers. This is now becoming standard practice in the field and it allowed us the flexibility to test multiple classifiers (we have presented only the classifiers that rank highest across our performance benchmarks). Hence, prioritizing validation studies based on published data has allowed us to obtain sufficient testing, optimization and validation to meet the primary goal of this task. The delay with the IHC studies also allowed us to be somewhat more productive and successful with some aspects of Tasks 1 and 2 than we had initially anticipated. We are currently preparing the results from the two clinical studies for publication.

ACCOMPLISHMENTS (for this one-year reporting period)

- Continued processing of specimens received from Edinburgh
- Obtained corresponding gene expression microarray data
- Continued improving electronic data storage and annotation infrastructure
- Updated data contained within the data storage and annotation infrastructure
- Built, tested and optimized classifier of TAM responsiveness using data stored in this electronic infrastructure
- Validated classifier of TAM responsiveness
- Built, tested, and optimized initial classifier of AI responsiveness
- Used data from this CTR to support additional (successful) applications for research funding to broaden the breast cancer research directions represented in this CTR
- Built, validated, and published new data analysis methods
    - cross phenotype normalization
    - differential dependency network method
    - multi-scale independent component analysis method for biomarker identification
- Began to identify new signaling network components associated with endocrine responsiveness

REPORTABLE OUTCOMES

Papers and Meeting Reports*

New Publications (for the present reporting period)

• Wang, Y., Miller, D.J. & Clarke, R. "Approaches to working in high dimensional data spaces: gene expression microarrays." *Br J Cancer,* 98: 1023-1028, 2008.

• Shajahan, A.N., Riggins, R.B. & Clarke, R. *"Apoptosis, cell death and breast cancer"* Chapter 8. In: *"Breast cancer: prognosis, treatment, and prevention."* Eds: Pasqualini, J.R., Informa Healthcare; New York, NY; pp137-156, 2008.

• Chen, L., Xuan, J., Wang, C., Shih, L-M., Wang, Y., Zhang, Z., Hoffman, E.P. & Clarke, R. "Knowledge-guided multi-scale independent component analysis for biomarker identification." *BMC Bioinformatics,* 9:416 (16 pages as published on-line), 2008

• Riggins, R.B., Lan, P.-J., Klimach, U., Zwart, A., Cavalli, L.R., Haddad, B.R, Chen, L., Xuan, J., Ethier, S.P. & Clarke, R. "ERRγ mediates Tamoxifen resistance in a novel model of invasive lobular breast carcinoma." *Cancer Res,* 68: 8908-8917, 2008.

• Zhang, B., Li, H., Riggins, R.B., Zhan, M., Xuan, J., Zhang, Z., Hoffman, E.P., Clarke, R. & Wang, Y. "Differential dependency network analysis to identify condition-specific topological changes in biological networks." *Bioinformatics,* 25: 526-532, 2009

• Clarke, R., Shajahan, A.N., Riggins, R.B., Cho, Y., Crawford, A., Xuan, J., Wang, Y., Zwart, A., Nehra, R. & Liu, M.C. "Gene network signaling in hormone responsiveness modifies apoptosis and autophagy in breast cancer cells." *J Steroid Biochem Mol Biol,* 114: 8-20, 2009

• Cavalli, L.R., Riggins, R.B., Wang, A., Clarke, R.*, & Haddad, B.R. "Frequent loss of heterozygosity at the interferon regulatory factor-1 (IRF1) gene locus in breast cancer" Breast Cancer Res Treat, in press, 2009.

• Xuan, J., Wang, Y., Hoffman, E. & Clarke, R. "Cross phenotype normalization of microarray data." Front Biosci, in press, 2009.

*We include in the appendix reprints of those papers that are already published and for which we have proofs or reprints. We do not list here or include in the appendices any published abstracts, but can do so if requested. Several other manuscripts also are submitted and in preparation – these will be provided to DOD in the future.

Comment on Subcontracts: Please also note that the majority of our publications here and in prior years include

coauthors from one or both of our subcontracts. Thus, our interdisciplinary and multi-institutional program worked very effectively and collaboratively; this should further be apparent in the development of new informatics methods (Virginia Polytechnic and State University subcontract) and the large number of high quality breast tumor specimens we obtained from the University of Edinburgh.

CONCLUSIONS

We have completed almost all of the work as proposed in the original application. We made significant progress on the research infrastructure goals and in the development and optimization of the methods needed for data analysis. We also have completed and published all of the data presented as preliminary data in the initial application and many of the results presented in prior reports. The clinical studies were held up by an unexpectedly long delay in obtaining final approval for our existing protocols - as noted by previous reviewers of our annual reports, this delay adversely affected the prospective study. Consistent with the recommendation of these prior reviewers, it was necessary to request a one-year no cost extension. This report is being submitted at the end of this period. We now present a TAM-specific and an AI-specific classifier. The former performs better than comparable classifiers, and it has been adequately validated in an independent data set. We are committed to continuing to follow the performance of patients on the prospective study and hope to obtain additional data to strengthen the performance of classifiers built on the prospectively accrued data. We have also begun to obtain important new insights into the biology of breast cancer in the context of its endocrine responsiveness. We have also begun to identify new individual biomarkers and potential targets for drug discovery. Thus, we have also positioned the work from this CTR to enable us to pursue aggressively the long terms goals of this research program. Overall, we are excited about the new knowledge gained from this study, the potential to move the research forward and make an impact on breast cancer to the benefit of patients. At the same time, we have successfully opened up new area of breast cancer research that has already attracted additional peer-reviewed support.

We wish to conclude by expressing our thanks to the DOD for its support of this important research program. We also wish to extend our thanks and admiration to those consumers who have, and who continue, to work tirelessly in support the DOD Breast Cancer Research Program that makes possible the work we have reported here and that performed by many of our peers.

REFERENCES

1. Early Breast Cancer Trialists Collaborative Group: Systemic treatment of early breast cancer by hormonal, cytotoxic, or immune therapy. Lancet, *399*: 1-15, 1992.

2. Early Breast Cancer Trialists' Collaborative Group. Tamoxifen for early breast cancer: an overview of the randomized trials. Lancet, *351*: 1451-1467, 1998.

3. Bonneterre, J., Buzdar, A., Nabholtz, J. M., Robertson, J. F., Thurlimann, B., von Euler, M., Sahmoud, T., Webster, A., and Steinberg, M. Anastrozole is superior to tamoxifen as first-line therapy in hormone receptor positive advanced breast carcinoma. Cancer, *92*: 2247-2258, 2001.

4. Bartek, J., Iggo, R., Gannon, J., and Lane, D. P. Genetic and immunochemical analysis of mutant p53 in human breast cancer cell lines. Oncogene, *5*: 893-899, 1990.

5. Winer, E. P., Hudis, C., Burstein, H. J., Chlebowski, R. T., Ingle, J. N., Edge, S. B., Mamounas, E. P., Gralow, J., Goldstein, L. J., Pritchard, K. I., Braun, S., Cobleigh, M. A., Langer, A. S., Perotti, J., Powles, T. J., Whelan, T. J., and Browman, G. P. American Society of Clinical Oncology technology assessment on the use of aromatase inhibitors as adjuvant therapy for women with hormone receptor-positive breast cancer: status report 2002. J Clin Oncol, *20*: 3317-3327, 2002.

6. Ellis, M. J., Coop, A., Singh, B., Mauriac, L., Llombert-Cussac, A., Janicke, F., Miller, W. R., Evans, D. B., Dugan, M., Brady, C., Quebe-Fehling, E., and Borgs, M. Letrozole is more effective neoadjuvant endocrine therapy than tamoxifen for ErbB-1- and/or ErbB-2-positive, estrogen receptor- positive primary breast cancer: evidence from a phase III randomized trial. J Clin Oncol, *19*: 3808-3816, 2001.

7. Miller, W. R., Anderson, T. J., and Dixon, J. M. Anti-tumor effects of letrozole. Cancer Invest, *20 Suppl 2*: 15-21, 2002.

8. Smith, I. E. and Dowsett, M. Aromatase inhibitors in breast cancer. N Engl J Med, *348*: 2431-2442, 2003.

9. Baum, M., Buzdar, A. U., Cuzick, J., Forbes, J., Houghton, J. H., Klijn, J. G., Sahmoud, T., and ATAC Trialists Group Anastrozole alone or in combination with tamoxifen versus tamoxifen alone for adjuvant treatment of postmenopausal women with early breast cancer: first results of the ATAC randomised trial. Lancet, *359*: 2131-2139, 2002.

10. Clarke, R., Leonessa, F., Welch, J. N., and Skaar, T. C. Cellular and molecular pharmacology of antiestrogen action and resistance. Pharmacol Rev, *53*: 25-71, 2001.

11. Clarke, R. and Brünner, N. Acquired estrogen independence and antiestrogen resistance in breast cancer: estrogen receptor-driven phenotypes? Trends Endocrinol Metab, *7*: 25-35, 1996.

12. Clarke, R., Skaar, T. C., Bouker, K. B., Davis, N., Lee, Y. R., Welch, J. N., and Leonessa, F. Molecular and pharmacological aspects of antiestrogen resistance. J Steroid Biochem Mol Biol, *76*: 71-84, 2001.

13. Loi, S., Haibe-Kains, B., Desmedt, C., Wirapati, P., Lallemand, F., Tutt, A. M., Gillet, C., Ellis, P., Ryder, K., Reid, J. F., Daidone, M. G., Pierotti, M. A., Berns, E. M., Jansen, M. P., Foekens, J. A., Delorenzi, M., Bontempi, G., Piccart, M. J., and Sotiriou, C. Predicting prognosis using molecular profiling in estrogen receptor-positive breast cancer treated with tamoxifen. BMC Genomics, *9*: 239, 2008.

14. Clarke, R., Ressom, H. W., Wang, A., Xuan, J., Liu, M. C., Gehan, E. A., and Wang, Y. The properties of very high dimensional data spaces: implications for exploring gene and protein expression data. Nature Rev Cancer, *8*: 37-49, 2008.

15. Shao, W., Halachmi, S., and Brown, M. ERAP140, a conserved tissue-specific nuclear receptor coactivator. Mol Cell Biol, *22*: 3358-3372, 2002.

16. Mullighan, C. G., Su, X., Zhang, J., Radtke, I., Phillips, L. A., Miller, C. B., Ma, J., Liu, W., Cheng, C., Schulman, B. A., Harvey, R. C., Chen, I. M., Clifford, R. J., Carroll, W. L., Reaman, G., Bowman, W. P., Devidas, M., Gerhard, D. S., Yang, W., Relling, M. V., Shurtleff, S. A., Campana, D., Borowitz, M. J., Pui, C. H., Smith, M., Hunger, S. P., Willman, C. L., and Downing, J. R. Deletion of IKZF1 and prognosis in acute lymphoblastic leukemia. N.Engl.J Med., *360*: 470-480, 2009.

British Journal of Cancer (2008) 00, 000–000

www.bjcancer.com

Minireview

Approaches to working in high-dimensional data spaces: gene expression microarrays

Y Wang*[,1], DJ Miller[2] and R Clarke[3,4]

[1]Department of Electrical, Computer, and Biomedical Engineering, Virginia Polytechnic Institute and State University, Arlington, VA 22203, USA; [2]Department of Electrical Engineering, The Pennsylvania State University, University Park, PA 16802, USA; [3]Departments of Physiology & Biophysics, Lombardi Comprehensive Cancer Center, Georgetown University, Washington, DC 20057, USA; [4]Departments of Oncology, Lombardi Comprehensive Cancer Center, Georgetown University, Washington, DC 20057, USA

This review provides a focused summary of the implications of high-dimensional data spaces produced by gene expression microarrays for building better models of cancer diagnosis, prognosis, and therapeutics. We identify the unique challenges posed by high dimensionality to highlight methodological problems and discuss recent methods in predictive classification, unsupervised subclass discovery, and marker identification.
British Journal of Cancer (2008) **0,** 000–000. doi:10.1038/sj.bjc.6604207 www.bjcancer.com

Keywords: microarray; gene expression profiling; high dimensionality; data modelling and analysis

Gene expression microarrays provide a wealth of information on gene expression patterns and cancer pathways with potential for (1) cancer diagnosis, prognosis, and prediction of therapeutic responsiveness (Ramaswamy *et al*, 2001; Dupuy and Simon, 2007); (2) discovering new cancer subtypes (Golub *et al*, 1999; Lange *et al*, 2004); and (3) identifying cancer-associated (signalling) molecular markers and their complex interactions (Shedden *et al*, 2003; Ransohoff, 2004). However, achieving these biological/clinical objectives requires comprehensive analysis of microarray gene expression profiles that exist in high-dimensional data spaces, and relies critically on the functional capabilities and accuracy of the relevant analytical techniques (Allison *et al*, 2006). Cancer diagnosis/prognosis and therapeutic responsiveness prediction are all supervised classification/prediction problems (Duda *et al*, 2001). Analysing gene expression patterns representing patients that manifest heterogeneous clinical outcomes to discover cancer subgroups amounts to an unsupervised clustering problem (Duda *et al*, 2001). Identification of cancer-associated markers can be cast either as supervised feature/gene selection or as multiple testing, with thousands of candidate markers and a small subset of true ones (Ransohoff, 2004).

Although these analytical tasks fall neatly within statistical learning and pattern recognition (Jain *et al*, 2000), there is nothing conventional about these tasks for microarray data analysis. Unlike conventional pattern recognition that involves moderately dimensioned data, usually less than 100 features per sample and hundreds to thousands of samples, microarrays often involve over 10 000 features/gene per sample (n) with typically at most several hundred clinical samples. A rule of thumb is to have at least 10 training samples per feature dimension (Jain *et al*, 2000), whereas in microarrays this ratio is often closer to 0.01 samples per dimension (Allison *et al*, 2006). High feature dimensionality and paucity of microarray samples pose unique challenges for, and inspire novel developments in, predictive classification, cluster discovery, and marker identification methodologies.

A common subtask is feature selection. For predictive classification, only a subset of discriminatory genes is used to avoid overfitting, where a classifier is known 'too well' to fit even irreproducible 'noisy' training patterns and, thus, to achieve predictive accuracy that generalises well to unseen/test data. In unsupervised clustering in high dimensions, feature selection is likewise essential for discerning the underlying grouping tendency that may be 'buried' in a much lower-dimensional subspace – with many structurally irrelevant features yet small sample size, clustering algorithms are likely to identify false group structure. Lastly, a separate objective is to identify cancer-associated genes and their joint effects, rather than to simply build a predictive model for the disease.

Although feature selection is integral to each of these analytical tasks, an exhaustive search of all 2^n-1 possible feature subsets is prohibitive for a large n. Thus, practical feature selection techniques are of necessity heuristic, with an inherent accuracy/complexity trade off. Moreover, while multivariate analysis methods based on complex criterion functions may reveal subtle joint marker effects, they are also prone to overfitting (Lai *et al*, 2006). Additionally, high dimensionality compromises the ability to validate marker discovery, which requires accurately measuring true and false discovery rates (Ransohoff, 2005). These issues have prompted the development of a variety of novel statistical methods for estimating (or controlling for) false discoveries (Storey, 2003).

PREDICTIVE CLASSIFICATION

Performance of a predictive model depends on the interrelationship between sample size, data dimensionality, and model

*Correspondence: Dr Y Wang; E-mail: yuewang@vt.edu
Received 5 September 2007; revised 13 December 2007; accepted 3 January 2008

complexity. The accuracy of learned models tends to deteriorate in high dimensions, a phenomenon called the 'curse of dimensionality' (Duda *et al*, 2001). This phenomenon is illustrated for classification by an example by Trunk (1979). Consider two equally probable, normally distributed classes with common variance in each dimension. For the feature indexed by $n = 1,2,3\ldots$, class 1 has mean $1/n^{1/2}$ and class 2 has mean $-1/n^{1/2}$. Thus, each additional feature has some class discrimination power, albeit diminishing as n increases. Trunk evaluated error rates for the Bayes decision rule, applied as a function of n, when the variance is assumed known but the class means are estimated based on a finite data set. Trunk found that (1) the best test error was achieved using a finite number of features; (2) using an infinite number of features, test error degrades to the accuracy of random guessing; and (3) the optimal dimensionality increases with increasing sample size. These observations are consistent with the 'bias/variance dilemma' (Jain *et al*, 2000). Simple models may be biased but will have low variance. More complex models have greater representation power (low bias) but overfit to the particular training set (high variance). Thus, the large variance associated with using many features (including those with modest discrimination power) defeats any possible classification benefit derived from these features (Figure 1). With severe limits on available samples in microarray studies, complex models using high-feature dimensions will severely overfit greatly compromising classification performance. Computational learning theory provides distribution-free bounds on generalisation accuracy in terms of a classifier's capacity related to model complexity (Vapnik, 1998). Relevance of these bounds to the microarray domain is discussed by Aliferis *et al* (2006).

There are some strategies for mitigating the aforementioned problem. One is to fit the high-dimensional data, but using simple models that restrict complexity such as naive Bayes models that assume features are conditionally independent or even simpler models that *share* some parameters across classes (Novovicova *et al*, 1996). Another approach is to apply support vector machines (SVMs), which attempt to avoid overfitting by finding a linear discriminant function (or generalised linear discriminant) that maximises the margin (the minimum distance of any sample point to the decision boundary) (Vapnik, 1998). The number of free parameters in SVMs is not a function of the dimensionality, but instead is upper-bounded by the number of samples, which for microarrays is much smaller (Ramaswamy *et al*, 2001). However, whether using linear or nonlinear kernels, SVMs are not immune to the curse of dimensionality. Finally, some methods aim to reduce the amount of parameter learning to avoid overfitting, achieved by regularisation techniques modifying the training objective function or limiting the parameter learning cycles (Duda *et al*, 2001).

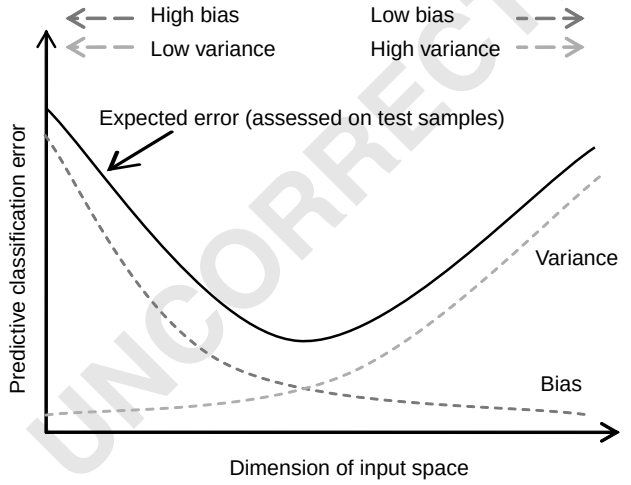


Figure 1 A demonstration of the bias/variance dilemma in predictive classification. Specifically, the error of model fitting can be decomposed into two components, bias (approximation error) and variance (estimation error). Added dimensions can degrade the prediction performance if the sample size is small relative to the dimensionality. For a fixed sample size in the high-dimensional data space, there is a trade off between the decreased approximation error and the increased estimation error.

Many microarray-based studies suggest that, irrespective of the classification method, feature selection is vital for achieving good generalisation performance (Statnikov *et al*, 2005). The vast number of feature subsets necessitates applying heuristic search techniques, with various accuracy/computation trade offs (Guyon and Elisseeff, 2003). Filtering methods apply knowledge of the class labels to evaluate the discrimination power either of individual genes (univariate) or collections of genes (multivariate), based on criteria such as signal-to-noise ratio, correlation measures, and mutual information, before classifier training. A recent study found that for small sample sizes, univariate methods faired comparably to multivariate methods, whose performance may be affected by overfitting (Lai *et al*, 2006).

Unlike filtering, wrapper-based approaches combine feature selection and classifier training, with the classifier learning algorithm repeatedly applied for different feature subsets and with the best subset chosen based on a specified criterion (Jain *et al*, 2000). These methods can improve predictive power by capturing higher order (and complex, nonlinear) joint feature effects. Perhaps the simplest example is the 'noisy XOR problem', for which two individual features and their linear combinations have no discrimination power, but a simple nonlinear combination is perfectly discriminating (Duda *et al*, 2001; Guyon and Elisseeff, 2003; Figure 2).

Wrapper algorithms, specified by the subset search method and the criterion for evaluating feature subsets, entail large computation in high dimensions, as the number of candidate spaces evaluated grows with the dimension. These algorithms include 'greedy' forward selection, with 'informative' features added starting from a null set. Other algorithms apply a backward search, which starts from the full space and then eliminates features. Floating (bidirectional) searches, which combine forward and backward steps, and more complex simulated annealing and genetic algorithms, can also be applied (Guyon and Elisseeff, 2003). Finally, there are methods that integrate classifier training and feature selection, such as decision trees, which essentially perform forward feature selection while growing a tree and backward elimination while pruning the tree (Duda *et al*, 2001). For evaluation criteria, either predictive accuracy on held-out test data (Statnikov *et al*, 2005), or criteria that can be evaluated solely on training data such as classifier margin or Bayesian model selection criteria (Guyon *et al*, 2002), can be used.

UNSUPERVISED CLUSTERING

In microarray data analysis, unsupervised clustering must be cautiously applied and may be unnecessary when samples come with appropriate and reliable supervising labels (Ramaswamy *et al*, 2001; Clarke *et al*, 2008). However, unsupervised clustering constitutes an important tool for discovering underlying cancer subtypes or gene modules (Frey and Dueck, 2007; Miller *et al*, 2008). Such exploration may suggest possible refinement to established cancer categories, where cancer subtypes manifest radically different clinical behaviour and may correspond to distinct biological pathways involving subtype-specific markers (Shedden *et al*, 2003). For example, prostate cancer can be an indolent cancer, remaining dormant throughout life, or an aggressive cancer leading to death. Similar issues arise in drug-resistance cases, where different cancer subtypes exhibit distinctive therapeutic responses (Golub *et al*, 1999).

NPG.BJC.9431-07REV

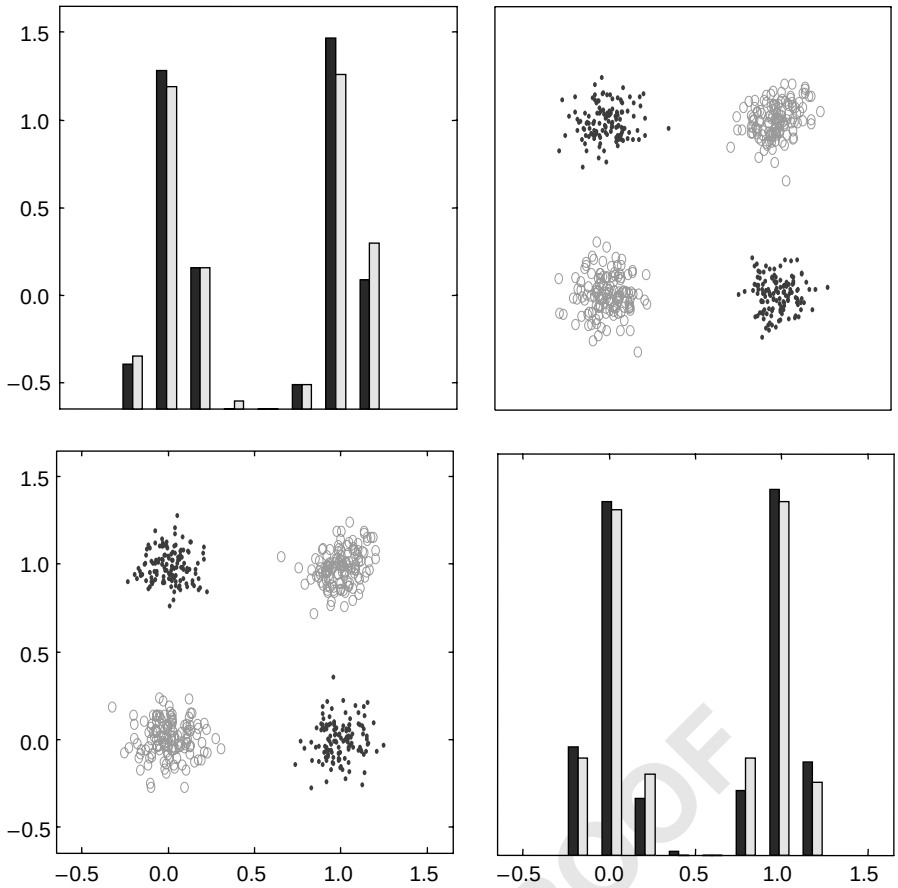


Figure 2 An example of XOR/chessboard-like joint effects. Although the classes consist of disjoint clusters, each variable has completely overlapping class conditional densities, that is, no marginal effect. In contrast, working together, the two variables provide good class separability.

Furthermore, when therapeutic responsiveness of patients is assessed based on interim growth or shrinkage of a tumour rather than the definitive clinical outcome, unsupervised clustering may be used to validate this supervision information, either to support it or to raise uncertainty about this 'ground-truth' if the correlation between the cluster labels and assessed responsiveness is weak. Moreover, trusted class labels on samples can be withheld during unsupervised clustering and subsequently used to validate the clustering methodology/assumptions. Strong correlation between clustering outcomes and known class labels supports the applicability of this clustering approach to other unlabelled microarray data (Golub *et al*, 1999).

While warranted in microarray data exploration, unsupervised clustering is extremely challenging in high dimensions with very few samples. Standard methods such as K-means and hierarchical clustering evaluate distances between data points using all (equally weighted) features. Thus, many noisy/irrelevant features will dominate the (much smaller set of) relevant features in determining how data points are partitioned, for example, many invariantly expressed genes used for microarray normalisation are irrelevant to classification or clustering. Rather than clustering samples using all genes, a practical alternative is to embed gene selection within unsupervised clustering – removal of noisy features improves clustering accuracy, which, in turn, guides a more accurate round of feature selection. Methods have been proposed along these lines (Xing and Karp, 2001; Graham and Miller, 2006), together with novel initialisation schemes (Frey and Dueck, 2007; Wang *et al*, 2007).

Another major challenge for clustering in high dimensions is estimating the number of clusters. Standard methods choose cluster number by best fitting the data while incurring least model complexity. However, under the widely used Bayesian information criterion (Duda *et al*, 2001), model complexity is linear in the number of parameters and quickly grows with each added feature. As many of these parameters model noisy/irrelevant features, their data fitting benefit is grossly outweighed by their contribution to model complexity, which leads to gross underestimation of the number of clusters. In a study by Graham and Miller (2006), a 'parsimonious' mixture model allows clusters to share distributions for noisy features, which enhances accuracy in estimating both the cluster parameters and the cluster number in high dimensions. Intrinsic to this modelling is identification of a distinct relevant feature subset specific to each sample cluster, that is, for the microarray domain, each subclass will have its own gene set, as has been conjectured by Shedden *et al* (2003); Ein-Dor *et al* (2005). Another strategy for identifying this cluster structure is top-down divisive clustering that explores and generates hierarchical mixtures in nested subspaces (Wang *et al*, 2007). By projecting high-dimensional data of a current cluster to multiple two-dimensional visualisation subspaces, the human gift for pattern recognition can be exploited to assess the current solution and assist further clustering refinement (Figure 3). Being more data-adaptive and process-transparent, human interaction may bring subjectivity, and thus must be carefully applied.

MARKER IDENTIFICATION

Marker identification aims to discover those genes and their complex interaction effects that have statistically significant correlations with cancer phenotypes. As it is currently largely unclear how molecular variants and their interactions determine

NPG.BJC.9431-07REV

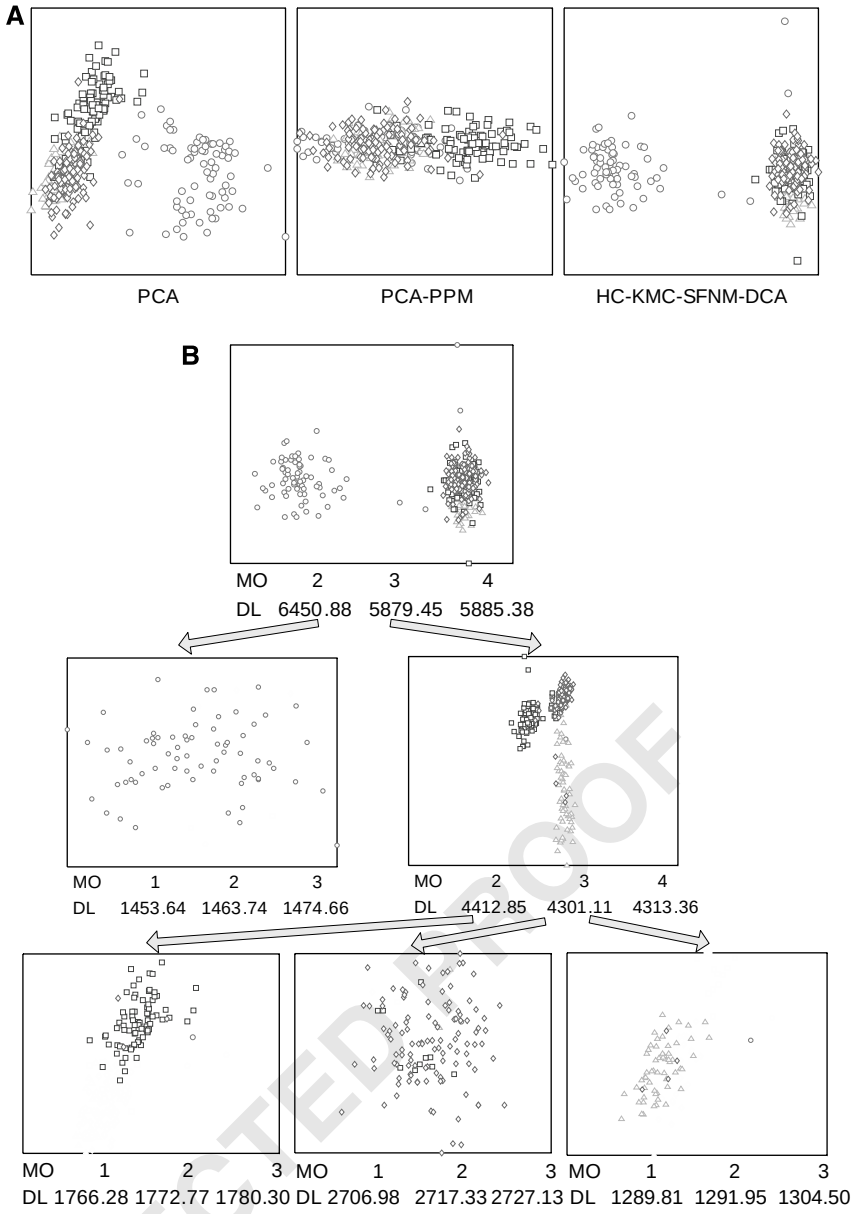


Figure 3 An example of coarse-to-fine top-down divisive unsupervised clustering using VISDA. (**A**) Multiple complementary visualisation subspaces derived from different data structure preserving projection principles. (**B**) Tree of phenotype with embedded model selection function, where MO refers to the model order (number of clusters) and DL refers to the description length (model complexity as a function of cluster number).

cancer pathogenesis and propensity, marker identification is valuable for improving understanding of the molecular mechanisms of cancers and for suggesting novel drug targets. Discovered markers may also define a subset of networked causal genes that regulate disease phenotype. A review of the current state of this effort is discussed by Aliferis *et al* (2006).

The objectives of feature selection for predictive classification and marker identification bear close resemblance. Although it is tempting to view these two problems as 'one and the same', this is often inappropriate. Inclusion of some true cancer markers in a feature set for cancer classification may provide negligible improvement in classification accuracy even though these markers are significantly associated with the cancer outcome of interest. A trivial example is where two markers are perfectly correlated, in which case only one of the two needs to be included in a predictive feature subset. A more interesting example is the one in which, even though two markers are only partially correlated, a classification model will not perceive any benefit from using both markers. This is illustrated below:

Let A and B take on one of the four possible discrete values and suppose the ground-truth statistics on class label C are $Prob[C=\text{'cancer'}|A=1]=1.0$; $Prob[C=\text{'cancer'}|A=i]=0.5$, $i=2,3,4$; $Prob[C=\text{'cancer'}|B=3]=1.0$; and $Prob[C=\text{'cancer'}|B=j]=0.5$, $j=1,2,4$. Suppose $Prob[A=1]=0.1$; $Prob[B=3]=0.7$; and $Prob[B=3|A=1]=0.5$. Thus, A and B are both informative about the disease (for one value), and these variables are only partially correlated. However, in a small training set, it is quite possible that each time $A=1$, $B=3$ also occurs, even though $Prob[B=3|A=1]$ is much less than one. In this case, while association-based marker discovery might include both A and B, classification-based marker discovery would only include B, because the training set suggests no predictive benefit from including A.

NPG.BJC.9431-07REV

More generally, whether predictive gene selection will include a gene that possesses some predictive benefit will depend on the sensitivity of the criterion function used. For example, a predictive model may achieve the same estimated classification error rate using several different feature subsets, even if there is a unique true marker subset, with greatest class discrimination power. Another limitation of predictive gene selection is that most classification models lack interpretability, that is, they do not allow easy discernment of the underlying interactions between the identified markers. The sole focus of most predictive feature selection techniques is on defeating the curse of dimensionality. Exceptions to this include decision trees (if not too large) and Bayesian networks (Duda *et al*, 2001).

Although association-based approaches may ultimately be found superior for identifying cancer markers and their interactions, these methods also have limitations. First, identifying marker interactions, particularly those involving markers with insignificant marginal effect, requires an exhaustive search over the full gene expression space. It is only practical to examine very low-order interactions, for example, '10 000 choose 2 or 3' possible interactions (Jain *et al*, 2000). Thus, higher-order interactions may get missed. One possible strategy is to first apply classification-based gene selection to significantly reduce the search space, followed by (exhaustive search) association-based marker identification. Second, it is difficult to evaluate and/or control inference accuracy for such testing, which involves numerous hypotheses. There is an inherent trade off between statistical power (true positive) and Type 1 error (false positive). Multiple testing for thousands of interacting genes at typical confidence levels leads to unacceptably large false positives. Family-wise error rate techniques can compensate, but conservatively toward minimising false positives and may have insufficient power. Other strategies improve inference accuracy through variance shrinkage that accounts for statistical dependencies between genes via computationally intensive permutation testing to accurately specify the null distribution.

To assess the true statistical significance of the implicated gene subset in multiple testing, one recent method is the randomisation–permutation test (Efron and Tibshirani, 2007). This method addresses the concern that a randomly selected gene subset may appear to possess significant association with the phenotype if only subjected to subject permutation testing. To assure that such false discoveries do not occur, a selected gene subset must, additionally, be subjected to a gene randomisation test, where the subject permutation test is to assess whether the implicated gene subset indeed has significant prediction power rather than 'by-chance' and the gene randomisation test assesses whether the implicated gene subset has significant prediction power as compared with that of any randomly selected gene subset of the same size.

An additional concern in marker identification is the impact of confounding variables (Ransohoff, 2005). A given data set may represent a biased sample with respect to factors such as patient age, gender, life style or with respect to sample handling, and expression levels for a putative marker may be more strongly associated with these confounding effects than with disease presence (Clarke *et al*, 2008). Although some confounding effects can be mitigated by careful study design or by explicitly accounting for these factors when performing marker identification, further research is needed to devise more effective methodologies for this purpose. Nevertheless, risk factors are not confounding effects to be discounted – there may be cancer-related gene–environment interactions that need to be identified. Finally, there are latent confounding sources due to biological multimodality. For complex phenotypes such as cancers, the presence of multiple, interrelated biological processes may obscure the true relationships between a gene subset and a specific outcome, creating spurious associations that appear statistically correct and yet may be false.

OUTCOME VALIDATION

In assessing the performance for any of three fundamental tasks, a validation procedure must be carefully designed, recognising limits on the accuracy of estimated performance, in particular for small sample size. In the study by Dupuy and Simon (2007), it was shown that, in more than 50% of a representative sample of past studies, inadequate statistical validation was performed. Clearly, classification accuracy must be assessed on labelled samples 'unseen' during training. However, single batch held-out test data are often precluded in microarray studies, as there will be insufficient samples for both accurate classifier training and accurate validation. The alternative is a sound cross-validation procedure, wherein all the data are used for both training and testing, but with held-out samples in a testing fold not used for any phase of classifier training, including feature selection and classifier design. Furthermore, performance (for either predictive classification or marker identification) depends on the threshold used to discriminate between categories. Most reported prediction accuracy rates are based on user-defined thresholds for a single operating point. A more meaningful estimate is the receiver operating characteristic curve obtained by using sensitivity (true positive rate) and specificity (true negative rate) acquired at a set of threshold values. The area under the curve gives a comprehensive figure-of-merit for prediction accuracy and can be shown to be a consistent but more sensitive measure than error rate for comparing classifiers, identifying performance differences between classifiers in cases where, evaluated solely by error rate, two classifiers would be deemed equivalent (Swets, 1988; Wang *et al*, 2006).

Unlike predictive classification assessment using labelled samples, validating unsupervised clustering requires alternative avenues when labels are not available. Synthetic data with constructed ground-truth may be used to assess the accuracy of a clustering or cluster number estimation algorithm. However, this approach will not validate that particular statistical assumptions are suitable for fitting molecular profiles from a given population. Alternatively, some form of cross-validation may be used to assess the 'stability' of clustering solutions (Lange *et al*, 2004). Stability analysis has been applied to clustering microarrays by Yeung *et al* (2001). Even when class labels are known, Dupuy and Simon (2007) suggest not to use them to select the gene space, as this will bias the clustering results.

It is even less likely to have ground-truth for validating marker identification. Synthetic data constructed from real microarray data can be used to assess a marker identification methodology, with class labels, markers, and interaction models handpicked and treated as ground truth. Importantly, 'reproducibility' of marker identification outcomes over multiple/bootstrap data sets may provide reasonable confidence (uncertainty assessment) on the discovered markers (Ransohoff, 2004).

Ultimately, discovered cancer markers or subtypes must be validated against definitive biomedical ground-truth. However, the cost of such validation demands a high degree of confidence in the knowledge extracted from microarray data by marker identification and clustering algorithms. Specifically, such knowledge extraction should not strongly depend on the particular random sample of data used or on variable aspects of the algorithms. Many clustering algorithms find only locally optimal solutions whose quality depends on the pseudorandomly chosen initial cluster parameter values (Frey and Dueck, 2007). Also, greedy sequential feature selection techniques are often 'unstable', giving results that may be highly dependent upon the particular training data used. There are two implications. First, whether synthetic data or real microarray data are used, extracted knowledge should be validated by assessing its reproducibility over multiple independently acquired data sets. Independent data sets are easily produced in the synthetic case, but at high cost in the case of real data. The second implication is that algorithms should be made as stable as possible to maximise the generalisation of their results. For marker discovery, one such strategy is to perform marker ranking multiple

NPG.BJC.9431-07REV

times, using bootstrap samples and/or *k*-fold cross-validation from the same data set, with the final, selected markers the ones with highest *average* ranking (and perhaps low variance/uncertainty). Nevertheless, the cost of increased stability in such approaches is an increase in computation.

ACKNOWLEDGEMENTS

This work was supported in part by the US National Institutes of Health under Grants CA109872, CA096483 and EB000830, and the US Department of Defense award BC030280.

REFERENCES

Aliferis CF, Statnikov A, Tsamardinos I (2006) Challenges in the analysis of mass-throughput data: a technical commentary from the statistical machine learning perspective. *Cancer Inform* **2:** 133 – 162

Allison DB, Cui X, Page GP, Sabripour M (2006) Microarray data analysis: from disarray to consolidation and consensus. *Nat Rev Genet* **7:** 55 – 65

Clarke R, Ressom HW, Wang A, Xuan J, Liu MC, Gehan EA, Wang Y (2008) The properties of high dimensional data spaces: implications for exploring gene and protein expression data. *Nat Rev Cancer* **8:** 37 – 49

Duda RO, Hart PE, Stork DG (2001) *Pattern Classification*, 2nd edn, New York: Wiley

Dupuy A, Simon RM (2007) Critical review of published microarray studies for cancer outcome and guidelines on statistical analysis and reporting. *J Nat Cancer Inst* **99:** 147 – 157

Efron B, Tibshirani R (2007) On testing the significance of sets of genes. *Ann Appl Stat* **1:** 107 – 129

Ein-Dor L, Kela I, Getz G, Givol G, Domany E (2005) Outcome signature genes in breast cancer: is there a unique set? *Bioinformatics* **21:** 171 – 178

Frey BJ, Dueck D (2007) Clustering by passing messages between data points. *Science* **315:** 972 – 976

Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD, Lander ES (1999) Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* **286:** 531 – 537

Graham MW, Miller DJ (2006) Unsupervised learning of parsimonious mixtures on large spaces with integrated feature and component selection. *IEEE Trans on Signal Process* **54:** 1289 – 1303

Guyon I, Elisseeff A (2003) An introduction to variable and feature selection. *J Mach Learn Res* **3:** 1157 – 1182

Guyon I, Weston J, Barnhill S, Vapnik V (2002) Gene selection for cancer classification using support vector machines. *Mach Learn* **46:** 389 – 422

Jain AK, Duin RPW, Mao J (2000) Statistical pattern recognition: a review. *IEEE Trans Pattern Anal Mach Intell* **22:** 4 – 37

Lai C, Reinders MJT, van't Veer LJ, Wessels LFA (2006) A comparison of univariate and multivariate gene selection techniques for classification of cancer datasets. *BMC Bioinformatics* **7:** 235 – 244

Lange T, Roth V, Braun ML, Buhmann JM (2004) Stability-based validation of clustering solutions. *Neural Comput* **16:** 1299 – 1323

Miller D, Wang Y, Kesidis G (2008) Emergent unsupervised clustering paradigms with potential application to bioinformatics. *Front Biosci* **13:** 677 – 690

Novovicova J, Pudil P, Kittler J (1996) Divergence-based feature selection for multimodal class densities. *IEEE Trans Pattern Anal Mach Intell* **18:** 218 – 223

Ramaswamy S, Tamayo P, Rifkin R, Mukherjee S, Yeang C-H, Angelo M, Ladd C, Reich M, Latulippe E, Mesirov JP, Poggio T, Gerald W, Loda M, Lander ES, Golub TR (2001) Multiclass cancer diagnosis using tumor gene expression signatures. *Proc Natl Acad Sci USA* **98:** 15149 – 15154

Ransohoff DF (2004) Rules of evidence for cancer molecular biomarker discovery and validation. *Nat Rev Cancer* **4:** 309 – 314

Ransohoff DF (2005) Bias as a threat to the validity of cancer molecular-marker research. *Nat Rev Cancer* **5:** 142 – 149

Shedden KA, Taylor JM, Giordano TJ, Kuick R, Misek DE, Rennert G, Schwartz DR, Gruber SB, Logsdon C, Simeone D, Kardia SL, Greenson JK, Cho KR, Beer DG, Fearon ER, Hanash S (2003) Accurate molecular classification of human cancers based on gene expression using a simple classifier with a pathological tree-based framework. *Am J Pathol* **163:** 1985 – 1995

Statnikov A, Aliferis CF, Tsamardinos I, Hardin D, Levy S (2005) A comprehensive evaluation of multicategory classification methods for microarray gene expression cancer diagnosis. *Bioinformatics* **21:** 631 – 643

Storey JD (2003) The positive false discovery rate: a Bayesian interpretation and the q-value. *Ann Stat* **31:** 2013 – 2035

Swets JA (1988) Measuring the accuracy of diagnostic systems. *Science* **240:** 1285 – 1293

Trunk GV (1979) A problem of dimensionality: a simple example. *IEEE Trans on Pattern Anal Mach Intell* **1:** 306 – 307

Vapnik VN (1998) *The Nature of Statistical Learning Theory*, Corrected 2nd printing edn, New York: Springer

Wang J, Li H, Zhu Y, Yousef M, Nebozhyn M, Showe M, Showe L, Xuan J, Clarke R, Wang Y (2007) VISDA: an open-source caBIG™ analytical tool for data clustering and beyond. *Bioinformatics* **23:** 2024 – 2027

Wang Z, Wang Y, Xuan J, Dong Y, Bakay M, Feng Y, Clarke R, Hoffman EP (2006) Optimized multilayer perceptrons for molecular classification and diagnosis using genomic data. *Bioinformatics* **22:** 755 – 761

Xing EP, Karp RM (2001) CLIFF: clustering of high-dimensional microarray data via iterative feature filtering using normalized cuts. *Bioinformatics* **17:** S306 – S315

Yeung KY, Haynor DR, Ruzzo WL (2001) Validating clustering for gene expression data. *Bioinformatics* **17:** 309 – 318

NPG.BJC.9431-07REV

8

Apoptosis, Cell Death, and Breast Cancer

AYESHA N. SHAJAHAN AND REBECCA B. RIGGINS

Department of Oncology, Lombardi Comprehensive Cancer Center, Georgetown University Medical Center, NW, Washington DC, U.S.A.

ROBERT CLARKE

Department of Oncology and Department of Physiology & Biophysics, Lombardi Comprehensive Cancer Center, Georgetown University Medical Center, NW, Washington DC, U.S.A.

INTRODUCTION

In 2007, approximately 213,000 new cases of invasive breast cancer will be diagnosed and over 41,000 American women will die of this disease (Jemal et al., 2007). Lifetime risk of developing breast cancer is modified by several factors related to development (e.g., weight at birth, age at menarche), reproductive life (e.g., parity, lactation, age at menopause), lifestyle (e.g., obesity, alcohol consumption), and inheritance (e.g., mutant BRCA1) (Ahlgren et al., 2004; de Jong et al., 2002; Feigelson et al., 2004; Hulka and Stark, 1995). Despite the importance of family history, the altered expression/function of tumor suppressor genes such as BRCA1/2 and TP53 do not account for the high prevalence of sporadic or non-BRCA familial breast cancers. Among mutant BRCA1/2 carriers, the timing of breast cancer onset and progress can vary substantially, but the factors responsible for these variations are not fully understood (Nathanson et al., 2001). The precise molecular events responsible for affecting disease progression remain unknown in both sporadic and inherited breast cancers, but those that affect a breast cancer cell's choice to proliferate, differentiate, or die are likely to be key factors in this process.

Randomized trials and large meta-analyses clearly show that all breast cancer patients derive a statistically significant survival benefit from chemotherapy and endocrine therapy (1998;2002b;2002a; Fisher et al., 1996; Mansour et al., 1998), Tamoxifen (TAM; antiestrogen), Paclitaxel (taxane), and Adriamycin (anthracycline) being among the most effective single agents. The survival benefit gained from current systemic therapies largely reflects the abilities of cytotoxic and endocrine agents to modify cell survival such that cells are driven down an irreversible cell death pathway (Fischer and Schulze-Osthoff, 2005). Nonetheless, advanced breast cancer largely remains an incurable disease, and new treatment regimens and schedules have led to only incremental decreases in breast cancer–related mortality. A better understanding of the factors that regulate breast cancer cell survival or death is central to improving breast cancer outcomes in women.

MOLECULAR MECHANISM OF APOPTOSIS

Apoptosis, a type of programmed cell death (PCD), is an essential feature of normal mammary gland function. Failure to undergo apoptosis can lead to the development of cancer

in breast epithelial cells (Green and Streuli, 2004). Cancer cells can evade apoptosis by modifying the signaling pathways that lead to apoptosis. Thus, significant effort is invested in the development of anticancer therapeutic agents that either selectively induce cell death in cancer cells or restore their apoptotic threshold (Meng et al., 2006).

The distinctive morphological and biochemical hallmarks of apoptosis include cell shrinkage, pyknosis (chromatin condensation), karyorhexis (nuclear fragmentation), membrane blebbing, and fragmentation of the cell into apoptotic bodies (Kerr et al., 1972). Phagocytic cells recognize and remove the apoptotic bodies, and thus avoid immune activation around the dying cell. Cell death by apoptosis requires the expenditure of ATP and the activation of proteases known as "caspases"(cysteine-dependent aspartate-specific proteases) (Wolf and Green, 1999). Caspases exist as latent zymogens that may be activated by autoactivation, transactivation, or proteolysis by other proteinases. In humans, over a dozen caspases have been identified (Hengartner, 2000); among these, caspases-3, -6, and -7 are called "executioner" caspases and they mediate their effect by the cleavage of specific cellular substrates. These executioner caspases are activated by the "initiator" caspases such as caspases-8, -9, and -10 (Denault and Salvesen, 2002; Riedl and Salvesen, 2007; Salvesen, 2002; Wolf and Green, 1999).

Overview of Extrinsic and Intrinsic Pathways

Many anticancer therapies activate caspases that cleave a number of different but selective substrates in the cytoplasm or nucleus, leading to many of the morphological features of an apoptotic cell death (Degterev et al., 2003). Caspase activation is initiated at the plasma membrane through death receptors either by an "extrinsic" pathway, or at the mitochondria by an "intrinsic" pathway (Fig. 1).

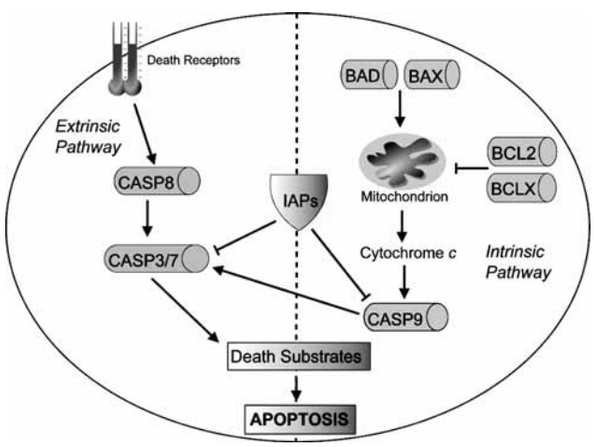


Figure 1 A simplified model of extrinsic and intrinsic signaling mechanisms involved in apoptotic cell death in the cell.

In normal tissue, apoptosis maintains homeostasis, and this process is tightly controlled at critical points of the signaling cascade (Green and Kroemer, 2004; Kroemer et al., 2007).

In the extrinsic or death receptor pathway, stimulation of death receptors of the tumor necrosis factor (TNF) receptor superfamily such as CD95 (APO-1/Fas) or TNF-related apoptosis-inducing ligand (TRAIL) receptors-1 and -2 result in the recruitment and oligomerization of the adapter molecule FADD (Fas-associating death domain-containing protein). The oligomerized FADD then localizes within the death-inducing signaling complex (Debatin and Krammer, 2004) followed by activation of initiator caspases-8, or -10 that contain death effector domains (Vandenabeele et al., 2006; Wolf and Green, 1999).

In the intrinsic or mitochondrial pathway, the executioner caspases are activated by caspase-9, which is activated by the adapter molecule apoptotic protease activating factor-1 (APAF-1) within a multiprotein complex called the "apoptosome." Activation of APAF-1 depends on both cytochrome *c* release from the intermembrane space of the mitochondria and ATP/dATP (Cain et al., 2002), leading to an increase in mitochondrial membrane potential (MMP) that is closely controlled by pro-(BAD, BAX) and antiapoptotic (BCL2) members of the BCL2 family of proteins (Decaudin et al., 1998; Green and Kroemer, 2004). A caspase-independent signal can also originate from within the mitochondria leading to irreversible loss of mitochondrial function; this can include the release of caspase-independent death effectors such as apoptosis-inducing factor (AIF) or endonuclease G (Cande et al., 2004; Kroemer and Martin, 2005; Li et al., 2001).

The intrinsic and extrinsic pathways converge at the executioner caspases. The executioner caspases selectively cleave their substrates in the primary sequence (always after an aspartate residue), and these target proteins can range from single polypeptide chain enzymes (poly ADP-ribose polymerse, PARP) to complex macromolecules (the lamin network) (Hengartner, 2000). PARP inactivation by caspase-specific cleavage, which forms an 89-kDa fragment, is a biochemical hallmark of apoptosis. Members of the heat shock protein (HSP) family, e.g., HSP70, can delay apoptosis by preventing the nuclear import of AIF (Ravagnan et al., 2001). The intrinsic drive for cancer cells to undergo apoptosis is held in check by inhibitor of apoptosis proteins (IAPs). Downstream of cytochrome *c* release, second mitochondrial activator of caspases/direct IAP binder with low pI (Smac/DIABLO) neutralize IAPs such as X-linked IAP (XIAP), survivin, and Apollon through their baculoviral inverted repeat (BIR) domains, and so indirectly promote caspase activation (Saelens et al., 2004; Vaux and Silke, 2003).

Alternative Death Pathways

Effectiveness of antineoplastic drugs can be assessed on the basis of their ability to induce apoptosis in tumor cells; however, it is now becoming evident that apoptosis may not be the only, or perhaps even the primary, mechanism of cell death in solid tumors (Brown and Attardi, 2005). Frequent failure to correlate apoptotic cell death with the effects of chemotherapeutic drugs on human tumors and cell lines (Brown and Attardi, 2005; Roninson et al., 2001) has prompted studies of other mechanisms of cell death.

Autophagy

Autophagy involves sequestration of cytosol and cytoplasmic organelles within double membranes called autophagosomes or autophagic vacuoles. The vesicular contents are broken down by pH-sensitive lysosomal hydrolases and the degradation products are recycled for use in macromolecular synthesis and/or bioenergetics (Kroemer and Jaattela, 2005). In general, autophagy is important in the developmental remodeling of cells, cellular adaptation to nutrient deprivation, and the elimination of damaged organelles (Edinger and Thompson, 2004; Edinger and Thompson, 2003; Klionsky and Emr, 2000). Paradoxically, autophagy can act both as a cell survival mechanism when extracellular nutrients or growth factors are limited and as an alternative cell death pathway to apoptosis (Jin, 2006) (Fig. 2). Removal of organelles such as mitochondria, which are apoptotic mediators, may protect the cell against apoptosis. Beclin-1/ATG6 (BECN1) is a key regulator of autophagy (Furuya et al., 2005), and monoallelic loss of the BECN1 locus is seen in over 40% of breast cancers (Liang et al., 1999). BCL2 antiapoptotic proteins can block autophagy by inhibiting BECN1 (Pattingre et al., 2005).

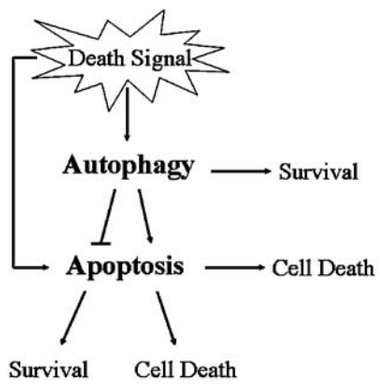


Figure 2 In response to a death signal, autophagy can prevent apoptosis to lead to survival or induce apoptosis to lead to cell death. The mechanism(s) that controls the balance between cell survival and cell death by autophagy, in response to a particular death signal, remains to be clarified.

Thus, antiapoptotic members of the BCL2 family may function as oncogenes not only by directly blocking apoptosis but also by blocking autophagy (Pattingre and Levine, 2006). Although the early events in autophagy are reversible, later events may share mechanism(s) with other death pathways. For example, cleavage of ATG5 by calpain (Yousefi et al., 2006) or upregulation of BID (Lamparska-Przybysz et al., 2006) can switch from autophagy to apoptosis.

Mitotic Catastrophe

Mitotic catastrophe is a type of cell death that occurs during or shortly after failed mitosis, which may occur following treatment with microtubule stabilizing or destabilizing agents or DNA damage (Mansilla et al., 2006; Roninson et al., 2001). The morphological alterations that occur during mitotic catastrophe are distinct from those that occur during apoptosis. These changes include multinucleation or the products of micronuclei because of faulty checkpoints, DNA structure checkpoints, or the spindle assembly checkpoint (also known as mitotic checkpoint) (Castedo et al., 2004; Roninson et al., 2001). The disruption of normal chromosome segregation of many chromosomes results in rapid cell death (Castedo et al., 2004). However, in the absence of cell death following mitotic catastrophe, the cell can divide asymmetrically, resulting in the generation of aneuploid daughter cells (Kops et al., 2005). Hence, mitotic catastrophe prevents irregular mitosis and, in turn, avoids aneuploidization that could lead to oncogenesis (Castedo et al., 2004; Kops et al., 2005).

Necrosis

A cell can undergo necrosis following physical damage or toxic insults when the intracellular level of ATP falls to a level incompatible with survival. The decision to undergo necrosis over apoptosis is dependent on the level of intracellular ATP, since apoptosis requires the presence of ATP, while necrosis results in ATP depletion (Nicotera et al., 1998). Morphologically, necrosis is identified by vacuolation of the cytoplasm, breakdown of the plasma membrane, and an induction of inflammation around the dying cell due to the release of cellular contents and proinflammatory molecules (Edinger and Thompson, 2004). The increase in cell volume (oncosis) during necrosis results in rupturing of the plasma membrane and the unorganized breakdown of swollen organelles (Kroemer et al., 2007). The ability of necrotic cells to promote local inflammation can support tumor growth (Vakkila and Lotze, 2004). While necrosis was previously thought to be a passive form of cell death, over the past several years the idea that cellular signaling pathways can specifically initiate necrosis has gained momentum (Proskuryakov et al., 2003).

Senescence

Cellular senescence was first identified as the state of permanent cell cycle arrest resulting from the replicative exhaustion of normal diploid cells in culture (Hayflick, 1965). Senescent cells appear static but are metabolically active, and appear large and flat with vacuoles and a large nucleus. In contrast to quiescent cells (reversible cell cycle arrest), senescent cells are unresponsive to mitogenic stimuli and are identified by cellular increase in β-galactosidase activity (Dimri et al., 1995). Thus, their inability to respond to serum or growth factors prevents immortalization and subsequent neoplastic transformation of the senescent cells. While apoptosis kills and eliminates potential cancer cells, cellular senescence irreversibly arrests cell growth (Campisi, 2001). Although the signaling mechanism for cellular senescence remains undetermined, DNA damage regulation by tumor suppressor genes such as TP53 and RB and epigenetic regulation of gene expression clearly play crucial roles (Campisi, 2001; Narita, 2007).

Telomeres are DNA-protein complexes that cap the end of linear eukaryotic chromosomes and range from 2 to 15 kb in humans (Martens et al., 1998). Telomeres prevent chromosomes from degradation, recombination, fusing with other chromosomes, and from being mistaken for DNA double-strand breaks. The de novo synthesis of telomeres is dependent on the enzyme telomerase, a reverse transcriptase (Cech, 2004). In most human cells, telomerase activity is gradually downregulated over time, resulting in successive telomere shortening that can ultimately limit their ability to proliferate; this is known as "cellular senescence," "mortality stage 1 (M1)," or "replicative senescence," since the maintenance of functional telomeres is crucial for continued proliferation. Inactivation of cell cycle checkpoint genes like TP53 can result in continued proliferation by bypassing the cell growth arrest in M1, eventually leading to critically short telomeres and massive cell death or "mortality stage 2 (M2)" or "crisis." Individual cells can evade M2 by maintaining their telomerase, resulting in immortal cancer cells that have been reported in 85% to 90% of human tumors (Kim et al., 1994). Thus, inhibition of telomerase activity may be a useful approach for mechanism-based anticancer therapy (reviewed in (Zimmermann and Martens, 2007).

ENDOCRINE THERAPIES

Endocrine therapy is often chosen as the first line of therapy in estrogen receptor α (ERα) and/or progesterone receptor (PR)-positive breast cancer patients because of its established efficacy and safety profile. However, endocrine resistance frequently arises. Two forms of endocrine resistance have been described: de novo resistance that is evident at the initial exposure to endocrine therapy or acquired resistance that arises over time after initiation of endocrine therapy. Absence of ERα expression is the most common de novo resistance mechanism, whereas a complete loss of ER expression is relatively uncommon in acquired resistance (Clarke et al., 2003; Moy and Goss, 2006). Improved knowledge of the mechanism of hormonal resistance, and the relationship between estrogen signaling and cell growth pathways, could provide the basis for combining signaling pathway inhibitors with endocrine therapies.

Antiestrogens, Aromatase Inhibitors, and Apoptosis

Endocrine therapy, administered as an antiestrogen (e.g., TAM or Faslodex) or an aromatase inhibitor (e.g., Letrozole or Anastrazole), is the least toxic and most effective means to manage hormone-dependent breast cancers. Antiestrogens, and TAM in particular, have been the "gold standard" first-line endocrine therapy for over 20 years. Newer antiestrogens such as Faslodex are also showing significantly improved activity relative to TAM and some aromatase inhibitors (Howell et al., 1995; Howell et al., 2002). Third generation aromatase inhibitors have emerged as viable alternatives to antiestrogens for first-line endocrine therapy; overall response rates are generally greater for aromatase inhibitors (Ferretti et al., 2006). Aromatase inhibitors also have a different mechanism of action and toxicity profile to antiestrogens, but whether this toxicity profile favors aromatase inhibitors over antiestrogens is controversial (Ferretti et al., 2006). Aromatase inhibitors can only be given as single agents to postmenopausal women or to women who do not have functioning ovaries; antiestrogens can be given irrespective of a patient's menopausal status.

The antiestrogen TAM, a nonsteroidal triphenylethylene derivative, has been used in the treatment of breast cancer for over 30 years (Litherland and Jackson, 1988). Besides being an effective means of treatment for women with ERα-positive tumors, TAM reduces the incidence of disease in healthy women at high risk of developing breast cancer (Cuzick et al., 2003). TAM acts by interacting with and blocking the action of estrogen on ERα within breast tumors and is classified as a selective ER modulator. In addition to the inhibitory effects of TAM on proliferation of breast ductal epithelium, TAM can act as an agonist to maintain bone density and to reduce serum cholesterol and triglyceride levels (Osborne et al., 2000). Despite its beneficial activities, prolonged administration of TAM can increase the incidence of endometrial cancer in some postmenopausal women (Wilking et al., 1997). The partial agonist properties of TAM have prompted the use of Faslodex to substitute for TAM as first- or

second-line endocrine treatment (Bundred and Howell, 2002; Howell, 2006). A steroidal analogue of 17β-estradiol, Faslodex (ICI 182,780; Fulvestrant) generally provides complete anatagonism without agonist effects, unlike TAM, which acts as a partial agonist. Faslodex prevents estrogen binding to ERα by inducing conformational changes and an eventual reduction of cellular ERα following polyubiquitination; Faslodex is classified as a selective downregulator of ER (SERD) {{74}}.

Antiestrogens primarily function through their ability to compete with estradiol (E2) for binding to ER and inducing growth arrest and cell death (Clarke et al., 2001). The consequences of occupying ER with an antiestrogen depend upon cellular context (i.e., expression of other proteins like ER coreguators), which ER is occupied (ERα, ERβ), and ligand structure (Clarke et al., 2001). The importance of ERα is established; that of ERβ is less clear (Speirs et al., 2004). Higher ERβ mRNA levels in resistant tumors have been reported (Arnold et al., 1995), but this cannot be causally linked to endocrine resistance because ERβ can also be associated with an aggressive phenotype (Dotzlaw et al., 1999; Speirs et al., 1999).

Aromatase inhibitors are often classed as type 1 (steroidal inactivator, e.g., Letrozole, Anastrozole), or type 2 (nonsteroidal inhibitor, e.g., exemestane). Third generation aromatase inhibitors are highly effective in blocking enzyme activity and exhibit notable specificity (Miller, 2004). These drugs act by blocking the ability of the P450 CYP19A1 gene product (aromatase) to convert androgen precursors to estrone or estradiol. Estradiol is the most potent estrogen and is found in high concentrations in breast tumors irrespective of menopausal status or the ER status of the tumor (Clarke et al., 2001; Clarke et al., 2003).

The extent of crossresistance among endocrine therapies is unclear. Since aromatase inhibitors can improve disease-free survival after two to three years of TAM as compared with a total of five years of TAM (Baum et al., 2003; Boccardo et al., 2001; Coombes et al., 2004; Jakesz et al., 2005; Thurlimann et al., 2005), it would seem that some TAM resistant tumors may retain sensitivity to an aromatase inhibitor. Second-and third-line responses to endocrine therapy have been widely documented, but lower response rates of shorter duration are usually observed with each successive line of treatment.

Both TAM and Faslodex can induce apoptosis in cells by inhibiting survival signaling mediated through ERα {{82, 80, 81}}. However, the precise mechanism for inducing apoptosis remains controversial. Administration of TAM results in activation of caspases-3, -8, -9 within 18 to 24 hours of drug treatment in rat mammary tumors (Mandlekar et al., 2000a). In addition, the effects of TAM can be mediated through an ERα-independent mechanism that results in an early activation of JNK1 followed by caspase activation (Mandlekar et al., 2000b). TAM can affect the level of proteins involved in cell growth including c-MYC (Kang et al., 1996), protein kinase C (Gelmann, 1996), and transforming growth factor β (TGFβ) (Perry et al., 1995) by ERα-independent mechanisms. However, since ER-negative tumors rarely respond to endocrine therapies (1998;2002a), these mechanisms are not likely to be responsible for significant apoptotic cell death in vivo.

About 25% of ER-positive/PR-positive tumors, 66% of ER-positive/PR-negative tumors and 55% of ER-negative/PR-positive tumors fail to respond to TAM (Clarke et al., 2003; Moy and Goss, 2006; Osborne and Schiff, 2003). Better predictors of endocrine responsiveness are clearly required. Many initially sensitive tumors become resistant (acquired resistance) (Clarke et al., 2001) and about one-third of all ER-positive breast tumors exhibit de novo endocrine resistance. The mechanisms of resistance to an antiestrogen remain unclear, reflecting a limited understanding of the signaling affecting cell proliferation, survival, and death and their hormonal regulation in breast cancer cells.

Of current interest is identification of the optimum choice and scheduling of antiestrogens and aromatase inhibitors. Evidence clearly shows improvements in overall response or disease-free survival for combined therapy (an aromatase inhibitor and an antiestrogen usually given sequentially) over single agent TAM (Baum et al., 2003; Boccardo et al., 2001; Coombes et al., 2004; Jakesz et al., 2005; Thurlimann et al., 2005). Recent data also imply that response rates are often greater to a first-line aromatase inhibitor than to TAM (Bonneterre et al., 2000; Mouridsen et al., 2001). However, the ability of aromatase inhibitors to induce a significant improvement in overall survival is uncertain. A recent meta-analysis failed to show an advantage for aromatase inhibitors with respect to overall survival, despite clear evidence favoring aromatase inhibitors in other end points (Ferretti et al., 2006). Thus, the optimum first-line endocrine therapy remains controversial for some women, as does the optimum choice and scheduling of combination endocrine therapy. Whichever way these controversies are eventually resolved, it is clear that both aromatase inhibitors and antiestrogens will remain as key modalities in the management of ER-positive breast cancers.

Understanding the signaling mechanism involved in antiestrogen-induced apoptosis is closely related to our knowledge of antiestrogen resistance. Through ongoing research in our laboratory, we have established interferon regulatory factor-1 (IRF-1) as a key node in a putative signaling network associated with responsiveness to endocrine therapy, where IRF-1 modifies ERα-mediated signaling to apoptosis (Bouker et al., 2004; Clarke et al., 2003; Gu et al., 2002). IRF-1 and a dominant negative IRF-1 (dnIRF-1) induce opposing effects on proliferation in vitro and tumorigenesis in vivo through regulation of

caspase-3/7 and caspase-8 activities (Bouker et al., 2005). While TP53-dependent apoptosis occurs in the breast (Tu et al., 2005), T47D cells express mutant TP53 and our data show that TP53 is not required for the proapoptotic actions of IRF-1 (Bouker et al., 2004; Bouker et al., 2005). Expression of IRF-1 is reduced in neoplastic versus normal human breast (Doherty et al., 2001), and there is an inverse correlation between IRF-1 and tumor grade (Connett et al., 2005). In a study of mostly ERα-positive breast tumors, nuclear expression of IRF-1 is negatively correlated with NFκB expression, suggesting their expression pattern to be consistent with other genes implicated in our signaling network for endocrine resistance (Zhu et al., 2006).

Upregulation of NFκB is associated with E2-independence (Clarkson and Watson, 1999; Nakshatri et al., 1997) and antiestrogen resistance (Gu et al., 2002; Riggins et al., 2005). The NFκB p50/p65 heterodimer complex comprises two homologous proteins encoded by different genes; the p50 product of its p105 precursor (NFκB1) and NFκB p65 (RELA). While the predominant form in breast cancer cell lines is NFκB (p50/p65), another member of the family (p52) also is expressed in some breast cancers (Cogswell et al., 2000). Perhaps reflecting its regulation by both E2 and growth factors (Biswas et al., 2000; Nakshatri et al., 1997), which are also involved in endocrine resistance (Clarke et al., 2001; Dickson and Lippman, 1995), normal mammary gland development appears to be dependent on NFκB (Clarkson and Watson, 1999). NFκB is maintained in the cytosol in an inactive state, e.g., complexed with members of the *IκB* family that either inhibit nuclear transport or block NFκB's nuclear translocation signal (Tam and Sen, 2001). Generally, activation proceeds by phosphorylation of IκB by the IKK kinase complex, which results in the ubiquitination and degradation of IκB (Yaron et al., 1998). Elevated NFκB activity arises during neoplastic transformation in both the rat (Kim et al., 2000) and mouse mammary gland (Tonko-Geymayer and Doppler, 2002).

Antiestrogens, Aromatase Inhibitors, Autophagy, and the Unfolded Protein Response

While apoptosis is clearly implicated (Bouker et al., 2004; Gaddy et al., 2004; Kyprianou et al., 1991), some of the apoptosis end points in prior studies may not distinguish among earlier events that are more closely linked to signaling initiated through autophagy. Autophagy does occur in response to endocrine therapy (Bursch et al., 1996; Inbal et al., 2002). However, given very recent advances in the understanding of cell death mechanisms, it is not clear if signaling to both apoptosis and autophagy are involved in regulating cell survival, and/or if the initial signaling involves autophagy but later events include signaling to apoptosis.

One cell signaling process that may integrate autophagy and apoptosis in this context is the unfolded protein response (UPR), a key component of the endoplasmic reticulum stress response (Ron, 2002) and an adaptive signaling pathway that allows cells to survive the accumulation of unfolded proteins in the endoplasmic reticulum (Zhang and Kaufman, 2006). Initially a mechanism for allowing cells to recover normal endoplasmic reticulum function, prolonged UPR can induce cell death. UPR is activated by three molecular sensors: IRE1α, ATF6, and PERK (DuRose et al., 2006). Splicing of X-box binding protein 1 (XBP1) by IRE1α is obligatory for IRE1α- and ATF6-induced UPR (DuRose et al., 2006; Yoshida et al., 2001). XBP1 is a transcription factor; the unspliced form, XBP1(U), has a molecular weight (Mr) of approximately 33 kDa and acts as a dominant negative (Lee et al., 2003; Sriburi et al., 2004). The spliced, active form, XBP1(S), has a Mr of approximately 54 kDa. A very recent study shows that the UPR (initiated by XBP1 splicing) can induce autophagy (Ogata et al., 2006). Whether this is a prosurvival or prodeath form of autophagy is unknown, since UPR can induce both prodeath and prosurvival outcomes (Feldman et al., 2005). In MCF7 and T47D breast cancer cells, we have shown that overexpression of XBP1(S) prevents antiestrogen-induced cell cycle arrest and cell death via the mitochondrial apoptotic pathway. Therefore, XBP1 may be a useful molecular target for the development of novel predictive and therapeutic strategies in breast cancer (Gomez et al., 2007, in press). Figure 3

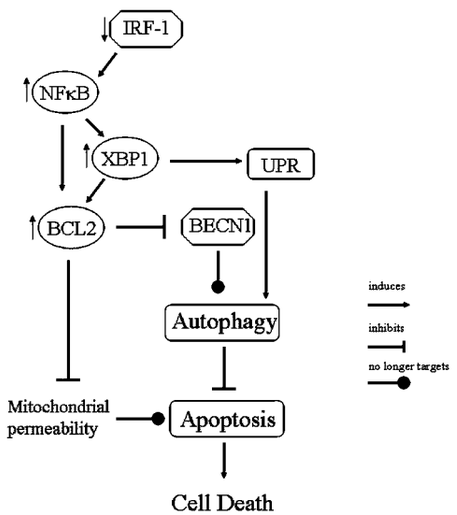


Figure 3 Some of the signaling mechanisms involved in endocrine resistance in breast cancer cells. *Abbreviations*: Interferon regulatory factor 1, IRF-1; nuclear factor kappa B, NFκB; X-box binding protein 1, XBP1; unfolded protein response, UPR; B-cell CLL/lymphoma 2, BCL2; beclin 1, BECN1.

shows some of the signaling mechanisms involved in endocrine resistance in breast cancer, based on the research done by others and our group.

CYTOTOXIC THERAPIES

Anthracyclines

Anthracyclines such as Doxorubicin are widely used in breast cancer treatment for the treatment of metastatic breast cancer. The mechanisms for the antineoplastic activities of doxorubucin are complex and include intercalation with DNA, direct cell membrane effects, initiation of DNA damage, apoptosis through inhibition of topoisomerase II, and the production of reactive oxygen species (Minotti et al., 2004). Despite its efficacy, Doxorubicin has several undesirable side effects, especially a cumulative cardiac toxicity (Singal and Iliskovic, 1998) (Table 1).

Doxorubicin-induced apoptosis occurs through a different signal transduction mechanism in nontransformed cells, such as endothelial cells and cardiomyocytes (H_2O_2-dependent), when compared with tumor cells that harbor a functional TP53 (Wang et al., 2004). Thus, targeted drug therapy could minimize cytotoxicity in normal cells and maximize cell death in cancer cells. The NF-κB/BCL2 pathway is a possible mechanism for tumor resistance to anthracyline-based chemotherapy. In breast tumor samples from patients treated with neoadjuvant Doxorubicin-based chemotherapy, nuclear localization of NF-κB is associated with expression of BCL2 and BAX. Moreover, the NF-κB/BCL2 pathway may be associated with a poor response to neoadjuvant Doxorubicin-based chemotherapy (Buchholz et al., 2005).

Alkylating agents

Cyclophosphamide (CPA) is an alkylating agent that alkylates DNA, forming DNA-DNA cross-links that result in an inhibition of DNA synthesis and cell death. CPA is a prodrug that is metabolically activated in the liver by cytochrome P450 enzymes. Activated CPA metabolites, e.g., hydroxyl-CPA, are transported via the bloodstream to both tumor and healthy tissues, where DNA and protein damage can occur (Moore, 1991). Cisplatin (cis-diaminedichloroplatinum II), another alkylating agent, exerts its cytotoxic effects by reacting with DNA to

Table 1

Drug	Primary target	Selected cell death signaling
Endocrine agents		
Faslodex, Tamoxifen	Estrogen receptor α	Caspase activation, BCL2 downregulation, JNK activation caspase activation, BAK upregulation
Anastrazole, Letrozole	CYP19A1 aromatase	
Anthracyclines		
Doxorubicin, Epirubicin	DNA intercalation, topoisomerase II	BCL2 family regulation, NFκB inhibition, TP53 activation
Alkylating agents		
Cisplatin, Cyclophosphamide	DNA crosslinking	Caspase activation, TP53 activation, cytochrome c release
Antimetabolites		
5-fluorouracil, Capecitabine	Thymidylate synthase	TP53 activation, thymineless death
Microtubule inhibitors		
Docetaxel, Paclitaxel	Microtubule stabilization	Caspase activation, phosphorylation of BCL2 and BCLX, JNK activation, CD95/FAS expression
Vinorelbine, Vincristine	Microtubule dissolution	TP53 activation, posttranslational modification of BCL2 family members
Signal transduction inhibitors		
Gefitinib	EGFR kinase activity	Phosphorylation of BAD, downregulation of BCL2
Trastuzumab, CH401	HER2 extracellular domain	Inhibition of PI3K/AKT, phosphorylation of BAD, JNK activation
Dasatinib, AZD0530	c-Src kinase activity	Inhibition of PI3K/AKT, downregulation of BCLX
Genasense	BCL2	Downregulation of BCL2
ABT-737, others	BCL2 and BCLX	Prevention of BCL2 and BCLX interaction with proapoptotic BAX and BAK
Bortezomib	26S proteasome	Inhibition of PI3K/AKT, JNK activation, sensitization to TRAIL
Bevacizumab	VEGFR2	DNA fragmentation, inhibition of PI3K/AKT

yield a variety of adducts, the most common adduct being an intrastrand cross-link between adjacent guanines (Perez, 1998). Cisplatin can be administered as a first-line therapy for metastatic breast carcinoma (Sledge Jr., et al., 1988) or in combination with other chemotherapeutic drugs (Kourousis et al., 1998; Nagourney et al., 2000).

Besides targeting genomic DNA, Cisplatin can also bind to mitochondrial DNA, interact with phospholipids and phosphatidylserine in membranes, disrupt the cytoskeleton, and affect the polymerization of actin (Jamieson and Lippard, 1999). Thus, Cisplatin interaction with proteins is also a possible mechanism for Cisplatin-induced apoptosis (Perez, 1998). Cisplatin can activate caspases by stabilizing TP53 and releasing cytochrome *c* from mitochondria (Siddik, 2002). Cell death by both apoptosis and necrosis has been found in the same population of ovarian cancer cells treated with Cisplatin (Pestell et al., 2000). Studies in MCF7 breast cancer cell have shown that antisense BCL2 and Cisplatin combination therapy could potentially be useful in treating breast cancer overexpressing BCL2, perhaps by activating caspase-8 independent of TP53 status (Basma et al., 2005). High doses of Cisplatin (>312 μM) can damage molecules involved in cellular energy supply (such as ATP) or proteins involved in apoptosis (such as TP53, caspases, BCL2, and BAX), leading to necrotic cell death. Thus, dose of cisplatin or the context of the target cells could direct the mode of cell death either by a defective apoptotic program or by necrosis (Gonzalez et al., 2001). Mechanisms of resistance to Cisplatin include loss of damage recognition, overexpression of HER-2/neu, activation of the PI3K/AKT (also known as PI3K/PKB) pathway, loss of TP53 function, overexpression of BCL2, and interference in caspase activation (Siddik, 2002).

Antimetabolites

Antimetabolites, particularly the fluoropyrimidine 5-fluorouracil (5-FU), are widely used as chemotherapeutic agents in the treatment of breast cancer. Within the cell, 5-FU is metabolized to 5-fluoro-deoxyuridine monophosphate (FdUMP) that interacts with and inhibits thymidylate synthase (TS) and prevents the formation of thymidine 5′-monophosphate (dTMP), thus inhibiting DNA synthesis (Chu et al., 2003; Longley et al., 2003). Thymidine phosphorylase (TP) mediates the conversion of 5-FU into fluorouridine diphosphate (FUDP) by transfer of a ribose phosphate from phosphoribosylpyrophosphate (PRPP) performed by orotic acid phosphoribosyltransferase (OPRTase). FUDP can be phosophorylated to FUTP and incorporated into RNA-by-RNA polymerase. FdUMP is converted to fluorodeoxyuridine triphosphate (FdUTP) and is directly incorporated into DNA (Chu et al., 2003; Longley et al., 2003).

The cytotoxic effect of 5-FU can occur through a process called thymineless death (Houghton et al., 1997). At the level of DNA synthesis, FdUMP acts as a competitive inhibitor of TS that results in a dTTP/dUMP cellular pool imbalance with subsequent DNA damage (Hernandez-Vargas et al., 2006; Parker and Cheng, 1990). Breast cancer patients with high pretreatment levels of TS protein show increased response to 5-FU-based chemotherapy (Nishimura et al., 1999). In addition, inhibition of RNA metabolism (Longley et al., 2003) and interference with polyamine metabolism (Zhang et al., 2003) could also contribute to the antiproliferative effects of 5-FU. In response to 5-FU treatment of MCF7 breast cancer cells, increased expression of TP53 target genes that are involved in cell cycle and apoptosis including CDKN1A (CIP1, WAF1), TP53INP, CD95/FAS, and BBC3/PUMA, along with significant repression of *MYC* (Hernandez-Vargas et al., 2006). Low doses of 5-FU (IC_{50}, 10 μM) result in cell cycle arrest, while high doses (IC_{80}, 500 μM) result in apoptosis in breast cancer cells (Hernandez-Vargas et al., 2006). Thus, the TP53 status of tumors and tissue concentration of 5-FU could be important determinants of drug efficacy in breast cancer treatment. Capecitabine is an example of a rationally designed cytotoxic drug that generates 5-FU preferentially in tumor cells by exploiting their higher activity of the activating enzyme TP compared with healthy tissues. The use of such targeted therapies increases efficacy and minimizes toxicity. Capecitabine has good activity and a favorable safety profile when used for the treatment of metastatic breast cancer (Yarden et al., 2004).

Microtubule Inhibitors

Microtubules form an integral part of the cytoskeleton and consist of α- and β- tubulin heterodimers. Taxanes (e.g., Taxol/Paclitaxel, Taxotere/Docetaxel) and vinca alkaloids (e.g., Vincristine, Vinorelbine) target microtubules and are often used to treat breast cancer. However, these drugs have very distinct effects on microtubule stability. While taxanes promote tubulin polymerization and stabilization (Schiff et al., 1979), vinca alkaloids inhibit tubulin polymerization (Johnson et al., 1960). Cells can die through either apoptotic or nonapoptotic pathways, but the precise signaling pathways involved are not completely elucidated. Q8

Paclitaxel induces aberrant mitotic spindle formation (Fuchs and Johnson, 1978) that results in G2-M cell cycle arrest followed by cell death. At concentrations lower than those required for microtubule disruption (2–4 nM), taxanes can inhibit cell growth, implying the existence of alternative mechanisms of action (Ganansia-Leymarie et al., 2003; Hernandez-Vargas et al., 2007). Several proapoptotic signaling mechanisms are implicated in response to antimicrotubule agents that appear independent of microtubule binding, such as phosphorylation of and

binding to BCL2/BCLX$_L$ antiapoptotic proteins (Haldar et al., 1997), activation of jun N-terminal kinase (JNK) (Wang et al., 1998) and RAF1 kinase (Blagosklonny et al., 1996). Other apoptosis-related proteins including TP53, CDKN1A, and BAX, and caspases are involved in apoptotic cell death process induced by taxanes (Ganansia-Leymarie et al., 2003). Phosphorylation of caveolin-1, a regulator of proapoptotic proteins, increases Paclitaxel sensitivity in MCF7 breast cancer cells by facilitating BCL2 phosphorylation and regulating the induction of CDKN1A (Shajahan et al., 2007). The death receptor CD95/FAS, a mediator of apoptosis, is induced after taxane therapy (Hernandez-Vargas et al., 2007). Treatment of breast cancer cells with low levels of Docetaxel (2–4 nM) triggers necrosis, while a higher concentration of the drug (100 nM) induces apoptosis (Hernandez-Vargas et al., 2007). Moreover, cell death by autophagy has been described as a form of PCD following treatment with taxanes in breast cancer cells (Gorka et al., 2005). Thus, both apoptotic and nonapoptotic pathways are likely responsible for cell death in response to taxanes.

Vinorelbine is an antimitotic drug that impairs chromosomal segregation during mitosis by blocking cell cycle at G2/M phase. Like other mitotic poisons, vinorelbine can induce apoptosis in cancer cells but an understanding of the signaling mechanism(s) of cell death mediated by vinca alkaloids is incomplete. Disruption of microtubules by vinca alkaloids can cause induction of TP53 and regulation of a number of proteins involved in apoptosis including CDKN1A, RAS/RAF, and PKC/PKA (Wang et al., 1999). As is the case with taxanes, BCL2 phosphorylation/inactivation plays a role in apoptosis induction (Haldar et al., 1995), resulting in a decrease in BCL2 inhibition of the proapoptotic protein BAX (Wang et al., 1999).

TARGETED INHIBITORS OF SIGNALING TRANSDUCTION PATHWAYS

In addition to the conventional cytotoxic and endocrine therapies discussed above, novel signal transduction inhibitors are emerging as valuable tools for inducing breast cancer cell death in vitro and in vivo. The precise molecular targets of these agents are varied, ranging from cell surface receptors to components of the proteasomal degradation pathway (recently reviewed in (Bremer et al., 2006). The purpose of these agents is to induce death in malignant cells while leaving normal tissue relatively unaffected, a goal that is potentially achievable if the signaling pathways contributing to malignant transformation are known. Given that a comprehensive analysis of all such agents as they pertain to breast cancer therapy is beyond the scope of this review, we will instead focus on three broad classes of molecules and their targeted therapies that have shown promise in the clinic.

The epidermal growth factor receptor (EGFR) and its family member HER2 are important mediators of breast cancer cell proliferation and survival [reviewed in (Badache and Gonsalves, 2006)]. Overexpression of EGFR and HER2 are associated with poor prognosis. While HER2 is amplified in approximately one-third of breast cancers, it is rare for either EGFR or HER2 to exhibit activating mutations (Diehl et al., 2007). Other receptors, such as insulin-like growth factor 1 receptor (IGF-1R), TGFβ receptor, and vascular endothelial growth factor receptor (VEGFR; see below) are additional targets that are receiving increased attention for their potential in the treatment of breast cancer.

Downstream of growth factor receptors, many intracellular signaling molecules coordinate the aberrant progrowth and prosurvival signals that lead to tumorigenesis. Other gene products are coupled to death receptors that transduce proapoptotic signals, while still others regulate cell proliferation and survival independent of growth factor receptors. The nonreceptor tyrosine kinase c-Src (Src) is overexpressed in approximately 70% of breast cancers, and the associated increase in its activity contributes significantly to tumor cell survival (recently reviewed in (Ishizawar and Parsons, 2004). The expression of intracellular signaling partners for TNF, FAS, TRAIL, and other cell surface death receptors may also be reduced, leaving cancer cells unable to respond to extrinsic apoptotic signals. More globally, deregulated expression of apoptotic gatekeepers like BCL2, or the ubiquitin-proteasome pathway, can disrupt the entire cell death program (Bremer et al., 2006).

In addition to the targeted therapies that directly disrupt cancer cell proliferation, others are designed to inhibit a tumor's ability to grow and spread by other means. Solid tumors such as breast cancer are particularly dependent on the formation of new blood vessels and capillary networks, both to supply fresh oxygen and nutrients and remove metabolic waste. The generation of new vascular tissue is known as angiogenesis; VEGF strongly induces angiogenesis by binding to the VEGFR2 expressed on endothelial cells. Many cancers, including those of the breast, overexpress VEGF, and agents that block angiogenesis via this signal transduction pathway have the capacity to induce apoptosis by starving the tumor [reviewed in (Schneider and Sledge, Jr., 2007).

Growth Factor Receptor Inhibitors

Preclinical studies of growth factor receptor inhibitors have revealed the potential of these agents to induce cell death by various means. The kinase inhibitor Gefitinib (ZD1839, Iressa) is an antagonist for EGFR but also has some activity toward HER2, and has been shown to

induce apoptosis in breast cancer cell lines through the mitochondrial or intrinsic apoptotic pathway. These effects arise either by decreased phosphorylation of proapoptotic BAD [reviewed in (Motoyama and Hynes, 2003)], or by down regulation of prosurvival BCL2 (Okubo et al., 2004). Okubo et al. also show that Gefitinib enhances apoptosis induced by the steroidal antiestrogen Faslodex in ER-positive breast cancer cell lines. Others have reported that Gefitinib can restore antiestrogen sensitivity in MCF7/HER2 xenografts that have become resistant to either Faslodex or estrogen deprivation (Massarweh et al., 2006). EGFR inhibition also has the potential to enhance or restore sensitivity to other chemotherapeutic agents. MCF7/ADR breast cancer cells are resistant to multiple drugs because of overexpression of the transport pump gp170/*MDR1,* and they have also been shown to express high levels of EGFR and its ligand TGFα (Ciardiello et al., 2002). However, Gefitinib is able to restore Paclitaxel- and Docetaxel-mediated apoptosis in MCF7/ADR cells, despite persistent expression of the transporter. Gefitinib may also prove useful in the treatment of high-risk women with ductal carcinoma in situ (DCIS); ER-positive and ER-negative DCIS grown as subcutaneous xenografts in nude mice are both sensitive to growth inhibition and the induction of apoptosis by this drug (Chan et al., 2002).

Inhibition of HER2 using the targeted monoclonal antibody Trastuzumab can also induce apoptosis in breast cancer cells through blocking HER2-mediated phosphoinositol 3-kinase (PI3K) and AKT activity. HER2-induced PI3K/AKT activity would normally promote cell survival by phosphorylating and inhibiting the proapoptotic functions of BAD (Badache and Gonsalves, 2006; Meric-Bernstam and Hung, 2006; Zhou and Hung, 2003). In addition, a novel inhibitory antibody targeted toward HER2 (CH401) has been shown to not only inhibit PI3K/AKT activities but also to directly induce apoptosis through modulation of JNK and p38 in gastric cancer cells (Hinoda et al., 2004). Whether this antibody will show similar activity toward HER2 in breast cancer remains to be determined.

Other growth factor receptor inhibitors have shown promise in both in vitro and preclinical studies. IGF-IR regulates cell proliferation, survival, and migration in multiple cancer models, using intracellular signal transduction pathways similar to those utilized by EGFR and HER2 (e.g., PI3K/AKT and RAS/MAPK) [recently reviewed in (Sachdev and Yee, 2007)]. However, unlike these receptors, IGF-IR signaling events cannot be induced by overexpression of the receptor and are exclusively ligand dependent. Therefore, strategies for inhibition of this pathway focus on blocking expression and/or function of the IGF-I ligand and its receptor. Humanized, single-chain antibodies directed against IGF-IR can induce downregulation of receptor expression in MCF7 breast cancer cells in vitro and grown as xenografts (Maloney et al., 2003; Sachdev et al., 2003). Over-expression of signal-inhibitory IGF binding proteins is also reported to inhibit IGF-IR signaling and cell growth while inducing breast cancer cell death [reviewed in (Maloney et al., 2003; Sachdev and Yee, 2007)].

Finally, TGFβ signaling represents another promising target for specific inhibition in breast cancer. It is well established that the type 1 TGFβ receptor can activate PI3K survival pathways and promote mammary epithelial cell survival (Muraoka-Cook et al., 2006; Yi et al., 2005). In the context of overexpressed HER2, TGFβ1 potently stimulates cell migration (Ueda et al., 2004). In addition, Arteaga et al. have shown that inhibitory antibodies to TGFβ2 can prevent the growth of established MCF7/LCC2 TAM-resistant xenografts (Arteaga et al., 1999). A number of approaches have been explored for inhibiting the TGFβ pathway, including small-molecule inhibitors, antisense oligonucleotides, and monoclonal antibody therapy (recently reviewed in (Lahn et al., 2005). One of these compounds, LY-580276, has been shown to significantly inhibit tumor development in MX1 xenografts, a model of ER-negative/PR-negative breast cancer. More recently, the TGFβ receptor kinase inhibitor SB-431542 has been demonstrated to abrogate epithelial-to-mesenchymal transition and wound healing in mouse mammary epithelial cells and the ER-negative/PR-negative MDA-MB-231 breast cancer cell line (Halder et al., 2005). Cell proliferation, as measured by [^{3}H] thymidine incorporation, is also inhibited under these conditions. However the molecular mechanism of cell death induced by these agents is unknown, and it is also important to consider that in many cell types TGFβ signaling has growth-inhibitory rather than growth-promoting effects.

Intracellular Signaling Inhibitors and Activators

Growth factor and other cell surface receptors can use diverse intracellular signaling partners to control cell growth and apoptosis. Receptors on the surface of a tumor cell can make for a relatively easy target for molecular inhibition. However, if the normal function and regulation of the intracellular signaling network(s) upon which these receptors rely is also disrupted, inhibiting only the receptor will be often ineffective. For example, altered downstream signaling may explain, in part, why the in vivo efficacy of EGFR and HER2 inhibitors has been mixed (Johnston, 2006). The use of agents designed to target intracellular signaling partners such as c-Src, BCL2, components of the proteasome, and transcription factors such as signal transducer and activator of transcription (STAT), STAT3 and STAT5, is likely to improve our ability to selectively induce apoptosis in breast cancer, perhaps in combination with receptor-targeted therapies.

c-Src's overexpression in a large percentage of breast cancers makes it an attractive target for therapeutic intervention [reviewed in (Ishizawar and Parsons, 2004)]. It is well established that EGFR and c-Src kinase activities synergistically promote breast cancer cell growth and survival (Biscardi et al., 2000). c-Src activity can also be regulated by physical and functional interactions with other proteins that include p130Cas (Burnham et al., 2000), and this has recently been shown to play a critical role in antiestrogen resistance in breast cancer (Riggins et al., 2006), in part through attenuating TAM-induced apoptosis. Despite these findings, there are currently no c-Src-specific inhibitors in clinical use for the treatment of breast cancer. Dasatinib (BMS-354825) is a dual inhibitor of c-Src and the Abl kinase which has recently been approved for use in chronic myelogenous leukemia (CML), and this agent has also recently been shown to inhibit the growth of "triple negative" breast cancer cell lines that lack expression of ER, PR, and EGFR, a subset of tumors that is often difficult to treat (Finn et al., 2007). AZD0530 is another dual c-Src/Abl inhibitor that has been shown to inhibit breast cancer cell growth and motility in vitro (Hiscox et al., 2006). In models of lung cancer and CML, Dasatinib can reduce AKT activity and expression of the prosurvival protein BCLX (Nam et al., 2007), although it is not yet known whether this also occurs in breast cancer.

Growth factor receptors and c-Src share the STAT family of proteins, specifically STAT3 and STAT5 as targets. These transcription factors regulate apoptosis in normal mammary gland development (Clarkson et al., 2006) and therefore are important players in breast cancer. However, their lack of enzymatic and ligand binding activities makes them difficult to target. Multiple alternative approaches have been considered for the targeted inhibition of STAT3 and STAT5, including small hairpin RNA, peptide mimetics that prevent binding to signaling partners, and small molecules such as sulindac and cyclooxygenase inhibitors [reviewed in (Desrivieres et al., 2006)].

The prosurvival protein BCL2 and components of the ubiquitin-proteasome system represent two signaling pathways that are essential to the proper regulation of apoptosis. Like STAT3/5 these also lack enzymatic activity, making them more difficult to target in the treatment of breast and other cancers. However, inhibition of BCL2 by the antisense oligonucleotide Genasense (oblimersen sodium, G3139) has been successful in improving the sensitivity of breast cancer preclinical models to apoptosis induced by conventional chemotherapies [reviewed in (Nahta and Esteva, 2003)], even in models of multiple-drug-resistant breast cancer (Lopes de Menezes et al., 2003). Genasense has also been studied in phase I clinical trials in combination with standard chemotherapeutics in many types of cancer including breast (Marshall et al., 2004) and hormone-refractory prostate cancer (Kim et al., 2007; Tolcher et al., 2004). More recently, small molecule inhibitors of *BCL2* function have been developed that are designed to disrupt its interaction with (and inhibition of) proapoptotic family members BAK and BAX. HA14-1 and YC137 are early-generation BCL2 inhibitors that have been shown to induce apoptosis alone and in combination with other drugs in breast cancer cell lines (Real et al., 2004; Witters et al., 2007). A newer BCL2/BCLX inhibitor (ABT-737) that has recently been developed may also prove useful in the sensitization of breast cancer to apoptosis induced by chemotherapeutic drugs (Dai and Grant, 2007).

The ubiquitin-dependent proteasome degradative pathway plays a critical role in maintaining appropriate cellular function by regulating the stability of signaling proteins. The classical pathway relies on conjugation of multiple ubiquitin moieties to the target protein, which is then degraded by the 26S proteasome. In contrast, monoubiquitination appears to regulate endocytosis and/or trafficking within the nucleus [reviewed in (Ohta and Fukuda, 2004)]. Both of these processes are active in breast cancer, and many important mediators of breast cancer cell proliferation and apoptosis are targets of the ubiquitin-proteasome pathway, including EGFR and ER (Marx et al., 2007). Consequently, there is considerable interest in studying the ability of proteasome inhibitors for the treatment of breast cancer (Dees and Orlowski, 2006). Bortezomib (PS-341, Velcade) is a small-molecule peptide mimetic that inhibits the 20S core of the proteasome that has been approved for the treatment of multiple myeloma (Voorhees and Orlowski, 2006). In SKBR3, MDA-MB-453, MCF7 and MCF7/HER2 breast cancer cell lines, Bortezomib can enhance apoptosis induced by the HER2 inhibitor Trastuzumab (Cardoso et al., 2006) and downregulate AKT activity, while stimulating apoptotic JNK activity, in combination with the multikinase inhibitor sorafenib in MDA-MB-231 breast cancer cells (Yu et al., 2006). Brooks et al. have shown that Bortezomib can sensitize some breast cancer cell lines to apoptosis induced by TRAIL, but this does not occur in all breast cancer models (Brooks et al., 2005). In a phase I study of Bortezomib in 12 patients with metastatic breast cancer, the drug showed minimal toxicity but was not efficacious in either stabilizing disease or improving outcome (Yang et al., 2006). It is likely that this, and other signal transduction inhibitors, will need to be used in combination with other therapeutics to be a viable treatment option for breast cancer.

Antiangiogenic Agents

Specific inhibition of VEGFR2 is a common approach for targeting angiogenesis in solid tumors, including those of the breast. VEGFR2 is expressed in the vasculature, where proliferation during the process of angiogenesis can be

stimulated in a paracrine manner by VEGF expressed by the tumor. Bevacizumab (Avastin) is a humanized monoclonal antibody directed against the VEGF ligand and the pursuit of similar angiogenesis inhibitors is an active focus of clinical breast cancer research [reviewed in (Schneider and Sledge Jr., 2007)]. Bevacizumab alone or in combination with Doxorubicin significantly inhibits VEGFR2 activation and angiogenic measures such as vascular permeability in women with inflammatory breast cancer (Wedam et al., 2006), and has been shown to significantly improve disease-free survival when combined with Paclitaxel in the treatment of metastatic breast cancer (Miller et al., 2005). Wedam et al. also showed that Bevacizumab induces apoptosis in vivo, as measured by terminal deoxynucleotidyl transferse-mediated dUTP nick-end labeling (TUNEL) staining (Wedam et al., 2006). In an in vitro model of lung tumorigenesis, Bevacizumab reduces AKT phosphorylation and activity (Inoue et al., 2007), although whether this occurs in the context of breast cancer is unknown.

SUMMARY AND CONCLUSIONS

We are now beginning to more clearly understand the molecular mechanisms of chemotherapy, endocrine therapy, and signal transduction inhibitor action and specifically how these compounds regulate tumor cell death. However, several key challenges remain. How we address these issues will directly affect our ability to be successful in the clinical management of breast cancer and in improving the outlook of women diagnosed with this disease.

One difficulty that is encountered is the heterogeneity among tumors and the multiple ways by which tumor cells can die in response to chemotherapeutic drugs. It is hoped that advancement of our knowledge of the molecular mechanisms of these drugs along with a greater understanding of tumorigenesis will enable more individualized treatment. In future, drug development and design must take into account the comprehensive cellular signaling mechanisms of inhibition of the target and the likely mechanisms by which resistance can develop to these drugs. In this regard, advances in biotechnology and bioinformatics will facilitate researchers and clinicians to assess high-throughput data gathered from DNA or proteomic arrays and tissue microarrays to enable molecular profiling of a patient's tumor. Thus, individual patients can be given a specific dose and type of chemotherapeutic drug that will increase efficacy and reduce unwanted toxicities.

A second challenge is redefining tumor cell death in the clinical setting. A reduction in tumor size is often considered evidence of a therapeutic agent's ability to induce cancer cell death. However, tumor shrinkage alone does not provide an understanding of the molecular mechanism(s) that control cell death in this context, and the signal transduction pathways that operate in vitro may not always be active in vivo. For many of the therapies discussed above, we do not yet understand whether the in vitro and in vivo cell death mechanisms of action are the same. This is particularly true for the newer, more specific signal transduction inhibitors. For example, when used in the neoadjuvant setting Trastuzumab can induce apoptosis in breast tumors, as measured by reduced AKT phosphorylation and increased cleavage of caspase-3 (Mohsin et al., 2005). In addition, Bevacizumab can induce DNA fragmentation in metastatic breast cancer (Wedam et al., 2006). However as more phase I/II clinical trials incorporate biomarker discovery into their design, we anticipate that these studies will generate important new data that clarify the in vivo apoptotic effects of these signal transduction inhibitors and other chemotherapeutic agents in breast cancer.

REFERENCES

Ahlgren M, Melbye M, Wohlfahrt J, Sorensen TI. Growth patterns and the risk of breast cancer in women. N Engl J Med 2004; 351:1619–1626.

Arnold SF, Obourn JD, Yudt MR, Carter, TH Notides, AC. In vivo and in vitro phosphorylation of the human estrogen receptor. J Steroid Biochem Mol Biol 1995; 52:159–171.

Arteaga CL, Koli KM, Dugger TC, Clarke R. Reversal of tamoxifen resistance of human breast carcinomas in vivo by neutralizing antibodies to transforming growth factor-beta. J Natl Cancer Inst 1999; 91:46–53.

Badache A, Gonsalves A. The ErbB2 signaling network as a target for breast cancer therapy. J Mammary Gland Biol Neoplasia 2006; 11:13–25.

Basma H, El-Refaey H, Sgagias MK, Cowan KH, Luo X, Cheng PW. BCL-2 antisense and cisplatin combination treatment of MCF-7 breast cancer cells with or without functional p53. J Biomed Sci 2005; 12:999–1011.

Baum M, Buzdar A, Cuzick J, Forbes J, Houghton J, Howell A, Sahmoud T. Anastrozole alone or in combination with tamoxifen versus tamoxifen alone for adjuvant treatment of postmenopausal women with early-stage breast cancer: results of the ATAC (Arimidex, Tamoxifen Alone or in Combination) trial efficacy and safety update analyses. Cancer 2003; 98:1802–1810.

Biscardi JS, Ishizawar RC, Silva CM, Parsons SJ. Tyrosine kinase signalling in breast cancer: epidermal growth factor receptor and c-Src interactions in breast cancer. Breast Cancer Res 2000; 2:203–210.

Biswas DK, Cruz AP, Gansberger E, Pardee AB. Epidermal growth factor-induced nuclear factor kappa B activation: A major pathway of cell-cycle progression in estrogen-receptor negative breast cancer cells. Proc Natl Acad Sci U S A 2000; 97:8542–8547.

Blagosklonny MV, Schulte T, Nguyen P, Trepel J, Neckers LM. Taxol-induced apoptosis and phosphorylation of Bcl-2 protein involves c-Raf-1 and represents a novel

c-Raf-1 signal transduction pathway. Cancer Res 1996; 56: 1851–1854.

Boccardo F, Rubagotti A, Amoroso D, Mesiti M, Romeo D, Caroti C, Farris A, Cruciani G, Villa E, Schieppati G, Mustacchi G. Sequential tamoxifen and aminoglutethimide versus tamoxifen alone in the adjuvant treatment of postmenopausal breast cancer patients: results of an Italian cooperative study. J Clin Oncol 2001; 19:4209–4215.

Bonneterre J, Buzdar A, Nabholtz JM, Robertson JF, Thurlimann B, von Euler M, Sahmoud T, Webster A, Steinberg M. Anastrozole is superior to tamoxifen as first-line therapy in hormone receptor positive advanced breast carcinoma. Cancer 2000; 92:2247–2258.

Bouker KB, Skaar TC, Fernandez DR, O'Brien KA, Riggins RB, Cao D, Clarke R. Interferon regulatory factor-1 mediates the proapoptotic but not cell cycle arrest effects of the steroidal antiestrogen ICI 182,780 (faslodex, fulvestrant). Cancer Res 2004; 64:4030–4039.

Bouker KB, Skaar TC, Riggins RB, Harburger DS, Fernandez DR, Zwart A, Wang A, Clarke R. Interferon regulatory factor-1 (IRF-1) exhibits tumor suppressor activities in breast cancer associated with caspase activation and induction of apoptosis. Carcinogenesis 2005; 26:1527–1535.

Bremer E, van Dam G, Kroesen BJ, de Leij L, Helfrich W. Targeted induction of apoptosis for cancer therapy: current progress and prospects. Trends Mol Med 2006; 12:382–393.

Brooks AD, Ramirez T, Toh U, Onksen J, Elliott PJ, Murphy WJ, Sayers TJ. The proteasome inhibitor bortezomib (Velcade) sensitizes some human tumor cells to Apo2L/TRAIL-mediated apoptosis. Ann N Y Acad Sci 2005; 1059: 160–167.

Brown JM, Attardi LD. The role of apoptosis in cancer development and treatment response. Nat Rev Cancer 2005; 5: 231–237.

Buchholz TA, Garg AK, Chakravarti N, Aggarwal BB, Esteva FJ, Kuerer HM, Singletary SE, Hortobagyi GN, Pusztai L, Cristofanilli M, Sahin AA. The nuclear transcription factor kappaB/bcl-2 pathway correlates with pathologic complete response to doxorubicin-based neoadjuvant chemotherapy in human breast cancer. Clin Cancer Res 2005; 11:8398–8402.

Bundred N, Howell A. Fulvestrant (Faslodex): current status in the therapy of breast cancer. Expert Rev Anticancer Ther 2002; 2:151–160.

Burnham MR, Bruce-Staskal PJ, Harte MT, Weidow CL, Ma A, Weed SA, Bouton AH. Regulation of c-SRC activity and function by the adapter protein CAS. Mol Cell Biol 2000; 20:5865–5878.

Bursch W, Ellinger A, Kienzl H, Torok L, Pandey S, Sikorska M, Walker R, Hermann RS. Active cell death induced by the anti-estrogens tamoxifen and ICI 164 384 in human mammary carcinoma cells (MCF-7) in culture: the role of autophagy. Carcinogenesis 1996; 17:1595–1607.

Cain K, Bratton SB, Cohen GM. The Apaf-1 apoptosome: a large caspase-activating complex. Biochimie 2002; 84:203–214.

Campisi J. Cellular senescence as a tumor-suppressor mechanism. Trends Cell Biol 2001; 11:S27–S31.

Cande C, Vahsen N, Kouranti I, Schmitt E, Daugas E, Spahr C, Luban J, Kroemer RT, Giordanetto F, Garrido C, Penninger JM, Kroemer G. AIF and cyclophilin A cooperate in apoptosis-associated chromatinolysis. Oncogene 2004; 23:1514–1521.

Cardoso F, Durbecq V, Laes JF, Badran B, Lagneaux L, Bex F, Desmedt C, Willard-Gallo K, Ross JS, Burny A, Piccart M, Sotiriou C. Bortezomib (PS-341, Velcade) increases the efficacy of trastuzumab (Herceptin) in HER-2-positive breast cancer cells in a synergistic manner. Mol Cancer Ther 2006; 5:3042–3051.

Castedo M, Perfettini JL, Roumier T, Andreau K, Medema R, Kroemer G. Cell death by mitotic catastrophe: a molecular definition. Oncogene 2004; 23:2825–2837.

Cech TR. Beginning to understand the end of the chromosome. Cell 2004; 116:273–279.

Chan KC, Knox WF, Gee JM, Morris J, Nicholson RI, Potten CS, Bundred NJ. Effect of epidermal growth factor receptor tyrosine kinase inhibition on epithelial proliferation in normal and premalignant breast. Cancer Res 2002; 62:122–128.

Chu E, Callender MA, Farrell MP, Schmitz JC. Thymidylate synthase inhibitors as anticancer agents: from bench to bedside. Cancer Chemother Pharmacol 2003; 52(suppl 1): S80–S89.

Ciardiello F, Caputo R, Borriello G, Del BD, Biroccio A, Zupi G, Bianco AR, Tortora G. ZD1839 (IRESSA), an EGFR-selective tyrosine kinase inhibitor, enhances taxane activity in bcl-2 overexpressing, multidrug-resistant MCF-7 ADR human breast cancer cells. Int J Cancer 2002; 98:463–469.

Clarke R, Leonessa F, Welch JN, Skaar TC. Cellular and molecular pharmacology of antiestrogen action and resistance. Pharmacol Rev 2001; 53:25–71.

Clarke R, Liu MC, Bouker KB, Gu Z, Lee RY, Zhu Y, Skaar TC, Gomez B, O'Brien K, Wang Y, Hilakivi-Clarke LA. Antiestrogen resistance in breast cancer and the role of estrogen receptor signaling. Oncogene 2003; 22:7316–7339.

Clarkson RW, Boland MP, Kritikou EA, Lee JM, Freeman TC, Tiffen PG, Watson CJ. The genes induced by signal transducer and activators of transcription (STAT)3 and STAT5 in mammary epithelial cells define the roles of these STATs in mammary development. Mol Endocrinol 2006; 20: 675–685.

Clarkson RW, Watson JC. NF-kappaB and apoptosis in mammary epithelial cells. J Mammary Gland Biol Neoplasia 1999; 4:165–175.

Cogswell PC, Guttridge DC, Funkhouser WK, Baldwin AS Jr. Selective activation of NF-kappa B subunits in human breast cancer: potential roles for NF-kappa B2/p52 and for Bcl-3. Oncogene 2000; 19:1123–1131.

Connett JM, Badri L, Giordano TJ, Connett WC, Doherty GM. Interferon regulatory factor 1 (IRF-1) and IRF-2 expression in breast cancer tissue microarrays. J Interferon Cytokine Res 2005; 25:587–594.

Coombes RC, Hall E, Gibson LJ, Paridaens R, Jassem J, Delozier T, Jones SE, Alvarez I, Bertelli G, Ortmann O, Coates AS, Bajetta E, Dodwell D, Coleman RE, Fallowfield LJ, Mickiewicz E, Andersen J, Lonning PE, Cocconi G, Stewart A, Stuart N, Snowdon CF, Carpentieri M, Massimini G, Bliss JM, d van V. A randomized trial of exemestane after two to three years of tamoxifen therapy in

postmenopausal women with primary breast cancer. N Engl J Med 2004; 350:1081–1092.

Cuzick J, Powles T, Veronesi U, Forbes J, Edwards R, Ashley A, Boyle P. Overview of the main outcomes in breast-cancer prevention trials. Lancet 2003; 361:296–300.

Dai Y, Grant S. Targeting multiple arms of the apoptotic regulatory machinery. Cancer Res 2007; 67:2908–2911.

Debatin KM, Krammer PH. Death receptors in chemotherapy and cancer. Oncogene 2004; 23:2950–2966.

Decaudin D, Marzo I, Brenner C, Kroemer G. Mitochondria in chemotherapy-induced apoptosis: a prospective novel target of cancer therapy (review). Int J Oncol 1998; 12:141–152.

Dees EC, Orlowski RZ. Targeting the ubiquitin-proteasome pathway in breast cancer therapy. Future Oncol 2006; 2: 121–135.

Degterev A, Boyce M, Yuan J. A decade of caspases. Oncogene 2003; 22:8543–8567.

de Jong MM, Nolte IM, Te Meerman GJ, van der Graaf WT, Oosterwijk JC, Kleibeuker JH, Schaapveld M, de Vries EG. Genes other than BRCA1 and BRCA2 involved in breast cancer susceptibility. J Med Genet 2002; 39:225–242.

Denault JB, Salvesen GS. Caspases: keys in the ignition of cell death. Chem Rev 2002; 102:4489–4500.

Desrivieres S, Kunz C, Barash I, Vafaizadeh V, Borghouts C, Groner B. The biological functions of the versatile transcription factors STAT3 and STAT5 and new strategies for their targeted inhibition. J Mammary Gland Biol Neoplasia 2006; 11:75–87.

Dickson RB, Lippman ME. Growth factors in breast cancer. Endocr Rev 1995; 16:559–589.

Diehl KM, Keller ET, Woods-Ignastoski KM. Why should we still care about oncogenes? Mol Cancer Ther 2007; 6:418–427.

Dimri GP, Lee X, Basile G, Acosta M, Scott G, Roskelley C, Medrano EE, Linskens M, Rubelj I, Pereira-Smith O, Peacocke M, Campisi J. A biomarker that identifies senescent human cells in culture and in aging skin in vivo. Proc Natl Acad Sci U S A 1995; 92:9363–9367.

Doherty GM, Boucher L, Sorenson K, Lowney J. Interferon regulatory factor expression in human breast cancer. Ann Surg 2001; 233:623–629.

Dotzlaw H, Leygue E, Watson PH, Murphy LC. Estrogen receptor-beta messenger RNA expression in human breast tumor biopsies: relationship to steroid receptor status and regulation by progestins. Cancer Res 1999; 59:529–532.

DuRose JB, Tam AB, Niwa M. Intrinsic capacities of molecular sensors of the unfolded protein response to sense alternate forms of endoplasmic reticulum stress. Mol Biol Cell 2006; 17:3095–3107.

Early Breast Cancer Trialists' Collaborative Group. Polychemotherapy for early breast cancer: an overview of the randomised trials. Lancet 1998; 352:930–942.

Edinger AL, Thompson CB. Death by design: apoptosis, necrosis and autophagy. Curr Opin Cell Biol 2004; 16:663–669.

Edinger AL, Thompson CB. Defective autophagy leads to cancer. Cancer Cell 2003; 4:422–424.

Feigelson HS, Jonas CR, Teras LR, Thun MJ, Calle EE. Weight gain, body mass index, hormone replacement therapy, and postmenopausal breast cancer in a large prospective study. Cancer Epidemiol Biomarkers Prev 2004; 13:220–224.

Feldman DE, Chauhan V, Koong AC. The unfolded protein response: a novel component of the hypoxic stress response in tumors. Mol Cancer Res 2005; 3:597–605.

Ferretti G, Bria E, Giannarelli D, Felici A, Papaldo P, Fabi A, Di CS, Ruggeri EM, Milella M, Ciccarese M, Cecere FL, Gelibter A, Nuzzo C, Cognetti F, Terzoli E, Carlini P. Second- and third-generation aromatase inhibitors as first-line endocrine therapy in postmenopausal metastatic breast cancer patients: a pooled analysis of the randomised trials. Br J Cancer 2006; 94:1789–1796.

Finn RS, Dering J, Ginther C, Wilson CA, Glaspy P, Tchekmedyian N, Slamon DJ. Dasatinib, an orally active small molecule inhibitor of both the src and abl kinases, selectively inhibits growth of basal-type/"triple-negative" breast cancer cell lines growing in vitro. Breast Cancer Res Treat 2007; 105(3):319–326. Epub 2007 Feb 1.

Fischer U, Schulze-Osthoff K. New approaches and therapeutics targeting apoptosis in disease. Pharmacol Rev 2005; 57: 187–215.

Fisher B, Dignam J, Mamounas EP, Costantino JP, Wickerham DL, Redmond C, Wolmark N, Dimitrov NV, Bowman DM, Glass AG, Atkins JN, Abramson N, Sutherland CM, Aron BS, Margolese RG. Sequential methotrexate and fluorouracil for the treatment of node-negative breast cancer patients with estrogen receptor-negative tumors: eight-year results from National Surgical Adjuvant Breast and Bowel Project (NSABP) B-13 and first report of findings from NSABP B-19 comparing methotrexate and fluorouracil with conventional cyclophosphamide, methotrexate, and fluorouracil. J Clin Oncol 1996; 14:1982–1992.

Fuchs DA, Johnson RK. Cytologic evidence that taxol, an antineoplastic agent from Taxus brevifolia, acts as a mitotic spindle poison. Cancer Treat Rep 1978; 62:1219–1222.

Furuya N, Yu J, Byfield M, Pattingre S, Levine B. The evolutionarily conserved domain of Beclin 1 is required for Vps34 binding, autophagy and tumor suppressor function. Autophagy 2005; 1:46–52.

Gaddy VT, Barrett JT, Delk JN, Kallab AM, Porter AG, Schoenlein PV. Mifepristone induces growth arrest, caspase activation, and apoptosis of estrogen receptor-expressing, antiestrogen-resistant breast cancer cells. Clin Cancer Res 2004; 10:5215–5225.

Ganansia-Leymarie V, Bischoff P, Bergerat JP, Holl V. Signal transduction pathways of taxanes-induced apoptosis. Curr Med Chem Anticancer Agents 2003; 3:291–306.

Gelmann EP. Tamoxifen induction of apoptosis in estrogen receptor-negative cancers: new tricks for an old dog? J Natl Cancer Inst 1996; 88:224–226.

Gomez BP, Riggins RB, Shajahan AN, Klimach U, Wang A, Crawford AC, Zhu Y, Zwart A, Wang M, Clarke R. Human X-Box binding protein-1 confers both estrogen independence and antiestrogen resistance in breastcancer cell lines. FASEB J 2007. In press.

Gonzalez VM, Fuertes MA, Alonso C, Perez JM. Is cisplatin-induced cell death always produced by apoptosis? Mol Pharmacol 2001; 59:657–663.

Gorka M, Daniewski WM, Gajkowska B, Lusakowska E, Godlewski MM, Motyl T. Autophagy is the dominant type of programmed cell death in breast cancer MCF-7 cells exposed to AGS 115 and EFDAC, new sesquiterpene analogs of paclitaxel. Anticancer Drugs 2005; 16:777–788.

Green DR, Kroemer G. The pathophysiology of mitochondrial cell death. Science 2004; 305:626–629.

Green KA, Streuli CH. Apoptosis regulation in the mammary gland. Cell Mol Life Sci 2004; 61:1867–1883.

Gu Z, Lee RY, Skaar TC, Bouker KB, Welch JN, Lu J, Liu A, Zhu Y, Davis N, Leonessa F, Brunner N, Wang Y, Clarke R. Association of interferon regulatory factor-1, nucleophosmin, nuclear factor-kappaB, and cyclic AMP response element binding with acquired resistance to Faslodex (ICI 182,780). Cancer Res 2002; 62:3428–3437.

Haldar S, Basu A, Croce CM. Bcl2 is the guardian of microtubule integrity. Cancer Res 1997; 57:229–233.

Haldar S, Jena N, Croce CM. Inactivation of Bcl-2 by phosphorylation. Proc Natl Acad Sci U S A 1995; 92: 4507–4511.

Halder SK, Beauchamp RD, Datta PK. A specific inhibitor of TGF-beta receptor kinase, SB-431542, as a potent antitumor agent for human cancers. Neoplasia 2005; 7:509–521.

Hayflick L. The limited in vitro lifetime of human diploid cell strains. Exp Cell Res 1965; 37:614–636.

Hengartner MO. The biochemistry of apoptosis. Nature 2000; 407:770–776.

Hernandez-Vargas H, Ballestar E, Carmona-Saez P, von KC, Banon-Rodriguez I, Esteller M, Moreno-Bueno G, Palacios J. Transcriptional profiling of MCF7 breast cancer cells in response to 5-Fluorouracil: relationship with cell cycle changes and apoptosis, and identification of novel targets of p53. Int J Cancer 2006; 119:1164–1175.

Hernandez-Vargas H, Palacios J, Moreno-Bueno G. Telling cells how to die: docetaxel therapy in cancer cell lines. Cell Cycle 2007; 6:780–783.

Hinoda Y, Sasaki S, Ishida T, Imai K. Monoclonal antibodies as effective therapeutic agents for solid tumors. Cancer Sci 2004; 95:621–625.

Hiscox S, Morgan L, Green TP, Barrow D, Gee J, Nicholson RI. Elevated Src activity promotes cellular invasion and motility in tamoxifen resistant breast cancer cells. Breast Cancer Res Treat 2006; 97:263–274.

Houghton JA, Harwood FG, Tillman DM. Thymineless death in colon carcinoma cells is mediated via fas signaling. Proc Natl Acad Sci U S A 1997; 94:8144–8149.

Howell A. Fulvestrant ('Faslodex'): current and future role in breast cancer management. Crit Rev Oncol Hematol 2006; 57:265–273.

Howell A, DeFriend D, Robertson J, Blamey R, Walton P. Response to a specific antioestrogen (ICI 182780) in tamoxifen-resistant breast cancer. Lancet 1995; 345:29–30.

Howell A, Robertson JF, Quaresma AJ, Aschermannova A, Mauriac L, Kleeberg UR, Vergote I, Erikstein B, Webster A, Morris C. Fulvestrant, formerly ICI 182,780, is as effective as anastrozole in postmenopausal women with advanced breast cancer progressing after prior endocrine treatment. J Clin Oncol 2002; 20:3396–3403.

Hulka BS, Stark AT. Breast cancer: cause and prevention. Lancet 1995; 346:883–887.

Inbal B, Bialik S, Sabanay I, Shani G, Kimchi A. DAP kinase and DRP-1 mediate membrane blebbing and the formation of autophagic vesicles during programmed cell death. J Cell Biol 2002; 157:455–468.

Inoue S, Hartman A, Branch CD, Bucana CD, Bekele BN, Stephens LC, Chada S, Ramesh R. mda-7 In combination with bevacizumab treatment produces a synergistic and complete inhibitory effect on lung tumor xenograft. Mol Ther 2007; 15:287–294.

Ishizawar R, Parsons SJ. c-Src and cooperating partners in human cancer. Cancer Cell 2004; 6:209–214.

Jakesz R, Jonat W, Gnant M, Mittlboeck M, Greil R, Tausch C, Hilfrich J, Kwasny W, Menzel C, Samonigg H, Seifert M, Gademann G, Kaufmann M, Wolfgang J. Switching of postmenopausal women with endocrine-responsive early breast cancer to anastrozole after 2 years' adjuvant tamoxifen: combined results of ABCSG trial 8 and ARNO 95 trial. Lancet 2005; 366:455–462.

Jamieson ER, Lippard SJ. Structure, recognition, and processing of Cisplatin-DNA adducts. Chem Rev 1999; 99: 2467–2498.

Jemal A, Siegel R, Ward E, Murray T, Xu J, Thun MJ. Cancer statistics, 2007. CA Cancer J Clin 2007; 57:43–66.

Jin S. Autophagy, mitochondrial quality control, and oncogenesis. Autophagy 2006; 2:80–84.

Johnson IS, Wright HF, Svoboda GH, Vlantis J. Antitumor principles derived from Vinca rosea Linn. I. Vincaleukoblastine and leurosine. Cancer Res 1960; 20:1016–1022.

Johnston SR. Targeting downstream effectors of epidermal growth factor receptor/HER2 in breast cancer with either farnesyltransferase inhibitors or mTOR antagonists. Int J Gynecol Cancer 2006; 16(Suppl 2):543–548.

Kang Y, Cortina R, Perry RR. Role of c-myc in tamoxifen-induced apoptosis estrogen-independent breast cancer cells. J Natl Cancer Inst 1996; 88:279–284.

Kerr JF, Wyllie AH, Currie AR. Apoptosis: a basic biological phenomenon with wide-ranging implications in tissue kinetics. Br J Cancer 1972; 26:239–257.

Kim DW, Sovak MA, Zanieski G, Nonet G, Romieu-Mourez R, Lau AW, Hafer LJ, Yaswen P, Stampfer M, Rogers AE, Russo J, Sonenshein GE. Activation of NF-kappaB/Rel occurs early during neoplastic transformation of mammary cells. Carcinogenesis 2000; 21:871–879.

Kim NW, Piatyszek MA, Prowse KR, Harley CB, West MD, Ho PL, Coviello GM, Wright WE, Weinrich SL, Shay JW. Specific association of human telomerase activity with immortal cells and cancer. Science 1994; 266:2011–2015.

Kim R, Emi M, Matsuura K, Tanabe K. Antisense and non-antisense effects of antisense Bcl-2 on multiple roles of Bcl-2 as a chemosensitizer in cancer therapy. Cancer Gene Ther 2007; 14:1–11.

Klionsky DJ, Emr SD. Autophagy as a regulated pathway of cellular degradation. Science 2000; 290:1717–1721.

Kops GJ, Weaver BA, Cleveland DW. On the road to cancer: aneuploidy and the mitotic checkpoint. Nat Rev Cancer 2005; 5:773–785.

Kourousis C, Kakolyris S, Androulakis N, P. Heras, J. Vlachonicolis, L.Vamvakas, M. Vlata, Hatzidaki D, Samonis G, Georgoulias V. Salvage chemotherapy with paclitaxel, vinorelbine, and cisplatin (PVC) in anthracycline-resistant advanced breast cancer. Am J Clin Oncol 1998; 21: 226–232.

Kroemer G, Galluzzi L, Brenner C. Mitochondrial membrane permeabilization in cell death. Physiol Rev 2007; 87:99–163.

Kroemer G, Jaattela M. Lysosomes and autophagy in cell death control. Nat Rev Cancer 2005; 5:886–897.

Kroemer G, Martin SJ. Caspase-independent cell death. Nat Med 2005; 11:725–730.

Kyprianou N, English HF, Davidson NE, Isaacs JT. Programmed cell death during regression of the MCF-7 human breast cancer following estrogen ablation. Cancer Res 1991; 51:162–166.

Lahn M, Kloeker S, Berry BS. TGF-beta inhibitors for the treatment of cancer. Expert Opin Investig Drugs 2005; 14:629–643.

Lamparska-Przybysz M, Gajkowska B, Motyl T. BID-deficient breast cancer MCF-7 cells as a model for the study of autophagy in cancer therapy. Autophagy 2006; 2:47–48.

Lee AH, Iwakoshi NN, Anderson KC, Glimcher LH. Proteasome inhibitors disrupt the unfolded protein response in myeloma cells. Proc Natl Acad Sci U S A 2003; 100:9946–9951.

Li LY, Luo X, Wang X. Endonuclease G is an apoptotic DNase when released from mitochondria. Nature 2001; 412:95–99.

Liang XH, Jackson S, Seaman M, Brown K, Kempkes B, Hibshoosh H, Levine B. Induction of autophagy and inhibition of tumorigenesis by beclin 1. Nature 1999; 402: 672–676.

Litherland S, Jackson IM. Antioestrogens in the management of hormone-dependent cancer. Cancer Treat Rev 1988; 15:183–194.

Longley DB, Harkin DP, Johnston PG. 5-fluorouracil: mechanisms of action and clinical strategies. Nat Rev Cancer 2003; 3:330–338.

Lopes de Menezes DE, Hu Y, Mayer LD. Combined treatment of Bcl-2 antisense oligodeoxynucleotides (G3139), p-glycoprotein inhibitor (PSC833), and sterically stabilized liposomal doxorubicin suppresses growth of drug-resistant growth of drug-resistant breast cancer in severely combined immunodeficient mice. J Exp Ther Oncol 2003; 3:72–82.

Maloney EK, McLaughlin JL, Dagdigian NE, Garrett LM, Connors KM, Zhou XM, Blattler WA, Chittenden T, Singh R. An anti-insulin-like growth factor I receptor antibody that is a potent inhibitor of cancer cell proliferation. Cancer Res 2003; 63:5073–5083.

Mandlekar S, Hebbar V, Christov K, Kong AN. Pharmacodynamics of tamoxifen and its 4-hydroxy and N-desmethyl metabolites: activation of caspases and induction of apoptosis in rat mammary tumors and in human breast cancer cell lines. Cancer Res 2000a; 60:6601–6606.

Mandlekar S, Yu R, Tan TH, Kong AN. Activation of caspase-3 and c-Jun NH2-terminal kinase-1 signaling pathways in tamoxifen-induced apoptosis of human breast cancer cells. Cancer Res 2000b; 60:5995–6000.

Mansilla S, Bataller M, Portugal J. Mitotic catastrophe as a consequence of chemotherapy. Anticancer Agents Med Chem 2006; 6:589–602.

Mansour EG, Gray R, Shatila AH, Tormey DC, Cooper MR, Osborne CK, Falkson G. Survival advantage of adjuvant chemotherapy in high-risk node-negative breast cancer: ten-year analysis–an intergroup study. J Clin Oncol 1998; 16:3486–3492.

Marshall J, Chen H, Yang D, Figueira M, Bouker KB, Ling Y, Lippman M, Frankel SR, Hayes DF. A phase I trial of a Bcl-2 antisense (G3139) and weekly docetaxel in patients with advanced breast cancer and other solid tumors. Ann Oncol 2004; 15:1274–1283.

Martens UM, Zijlmans JM, Poon SS, Dragowska W, Yui J, Chavez EA, Ward RK, Lansdorp PM. Short telomeres on human chromosome 17p. Nat Genet 1998; 18:76–80.

Marx C, Yau C, Banwait S, Zhou Y, Scott GK, Hann B, Park JW, Benz CC. Proteasome-regulated ERBB2 and estrogen receptor pathways in breast cancer. Mol Pharmacol 2007; 71: 1525–1534.

Massarweh S, Osborne CK, Jiang S, Wakeling AE, Rimawi M, Mohsin SK, Hilsenbeck S, Schiff R. Mechanisms of tumor regression and resistance to estrogen deprivation and fulvestrant in a model of estrogen receptor-positive, HER-2/neu-positive breast cancer. Cancer Res 2006; 66:8266–8273.

Meng XW, Lee SH, Kaufmann SH. Apoptosis in the treatment of cancer: a promise kept? Curr Opin Cell Biol 2006; 18: 668–676.

Meric-Bernstam F, Hung MC. Advances in targeting human epidermal growth factor receptor-2 signaling for cancer therapy. Clin Cancer Res 2006; 12:6326–6330.

Miller KD, Chap LI, Holmes FA, Cobleigh MA, Marcom PK, Fehrenbacher L, Dickler M, Overmoyer BA, Reimann JD, Sing AP, Langmuir V, Rugo HS. Randomized phase III trial of capecitabine compared with bevacizumab plus capecitabine in patients with previously treated metastatic breast cancer. J Clin Oncol 2005; 23:792–799.

Miller WR. Biological rationale for endocrine therapy in breast cancer. Best Pract Res Clin Endocrinol Metab 2004; 18: 1–32.

Minotti G, Menna P, Salvatorelli E, Cairo G, Gianni L. Anthracyclines: molecular advances and pharmacologic developments in antitumor activity and cardiotoxicity. Pharmacol Rev 2004; 56:185–229.

Mohsin SK, Weiss HL, Gutierrez MC, Chamness GC, Schiff R, Digiovanna MP, Wang CX, Hilsenbeck SG, Osborne CK, Allred DC, Elledge R, Chang JC. Neoadjuvant trastuzumab induces apoptosis in primary breast cancers. J Clin Oncol 2005; 23:2460–2468.

Moore MJ. Clinical pharmacokinetics of cyclophosphamide. Clin Pharmacokinet 1991; 20:194–208.

Motoyama AB, Hynes NE. BAD: a good therapeutic target? Breast Cancer Res 2003; 5:27–30.

Mouridsen H, Gershanovich M, Sun Y, Perez-Carrion R, Boni C, Monnier A, Apffelstaedt J, Smith R, Sleeboom HP, Janicke F, Pluzanska A, Dank M, Becquart D, Bapsy PP, Salminen E, Snyder R, Lassus M, Verbeek JA, Staffler B, Chaudri-Ross HA, Dugan M. Superior efficacy of letrozole versus tamoxifen as first-line therapy for postmenopausal women with advanced breast cancer: results of a phase III study of the International Letrozole Breast Cancer Group. J Clin Oncol 2001; 19:2596–2606.

Moy B, Goss PE. Estrogen receptor pathway: resistance to endocrine therapy and new therapeutic approaches. Clin Cancer Res 2006; 12:4790–4793.

Multi-agent chemotherapy for early breast cancer. Cochrane Database Syst Rev 2002a CD000487.

Muraoka-Cook RS, Shin I, Yi JY, Easterly E, Barcellos-Hoff MH, Yingling JM, Zent R, Arteaga CL. Activated type I TGFbeta receptor kinase enhances the survival of mammary epithelial cells and accelerates tumor progression. Oncogene 2006; 25:3408–3423.

Nagourney RA, Link JS, Blitzer JB, Forsthoff C, Evans SS. Gemcitabine plus cisplatin repeating doublet therapy in previously treated, relapsed breast cancer patients. J Clin Oncol 2000; 18:2245–2249.

Nahta R, Esteva FJ. Bcl-2 antisense oligonucleotides: a potential novel strategy for the treatment of breast cancer. Semin Oncol 2003; 30:143–149.

Nakshatri H, Bhat-Nakshatri P, Martin DA, Goulet RJ Jr., Sledge GW Jr. Constitutive activation of NF-kappaB during progression of breast cancer to hormone-independent growth. Mol Cell Biol 1997; 17:3629–3639.

Nam S, Williams A, Vultur A, List A, Bhalla K, Smith D, Lee FY, Jove R. Dasatinib (BMS-354825) inhibits Stat5 signaling associated with apoptosis in chronic myelogenous leukemia cells. Mol Cancer Ther 2007; 6:1400–1405.

Narita M. Cellular senescence and chromatin organisation. Br J Cancer 2007; 96:686–691.

Nathanson KL, Wooster R, Weber BL. Breast cancer genetics: what we know and what we need. Nat Med 2001; 7: 552–556.

Nicotera P, Leist M, Ferrando-May E. Intracellular ATP, a switch in the decision between apoptosis and necrosis. Toxicol Lett 1998; 102–103:139–142.

Nishimura R, Nagao K, Miyayama H, Matsuda M, Baba K, Matsuoka Y, Yamashita H, Fukuda M, Higuchi A, Satoh A, Mizumoto T, Hamamoto R. Thymidylate synthase levels as a therapeutic and prognostic predictor in breast cancer. Anticancer Res 1999; 19:5621–5626.

Ogata M, Hino S, Saito A, Morikawa K, Kondo S, Kanemoto S, Murakami T, Taniguchi M, Tanii I, Yoshinaga K, Shiosaka S, Hammarback JA, Urano F, Imaizumi K. Autophagy is activated for cell survival after endoplasmic reticulum stress. Mol Cell Biol 2006; 26:9220–9231.

Ohta T, Fukuda M. Ubiquitin and breast cancer. Oncogene 2004; 23:2079–2088.

Okubo S, Kurebayashi J, Otsuki T, Yamamoto Y, Tanaka K, Sonoo H. Additive antitumour effect of the epidermal growth factor receptor tyrosine kinase inhibitor gefitinib (Iressa, ZD1839) and the antioestrogen fulvestrant (Faslodex, ICI 182,780) in breast cancer cells. Br J Cancer 2004; 90: 236–244.

Osborne CK, Schiff R. Growth factor receptor cross-talk with estrogen receptor as a mechanism for tamoxifen resistance in breast cancer. Breast 2003; 12:362–367.

Osborne CK, Zhao H, Fuqua SA. Selective estrogen receptor modulators: structure, function, and clinical use. J Clin Oncol 2000; 18:3172–3186.

Parker WB, Cheng YC. Metabolism and mechanism of action of 5-fluorouracil. Pharmacol Ther 1990; 48:381–395.

Pattingre S, Levine B. Bcl-2 inhibition of autophagy: a new route to cancer? Cancer Res 2006; 66:2885–2888.

Pattingre S, Tassa A, Qu X, Garuti R, Liang XH, Mizushima N, Packer M, Schneider MD, Levine B. Bcl-2 antiapoptotic proteins inhibit Beclin 1-dependent autophagy. Cell 2005; 122:927–939.

Perez RP. Cellular and molecular determinants of cisplatin resistance. Eur J Cancer 1998; 34:1535–1542.

Perry RR, Kang Y, Greaves BR. Relationship between tamoxifen-induced transforming growth factor beta 1 expression, cytostasis and apoptosis in human breast cancer cells. Br J Cancer 1995; 72:1441–1446.

Pestell KE, Hobbs SM, Titley JC, Kelland LR, Walton, MI. Effect of p53 status on sensitivity to platinum complexes in a human ovarian cancer cell line. Mol Pharmacol 2000; 57:503–511.

Proskuryakov SY, Konoplyannikov AG, Gabai VL. Necrosis: a specific form of programmed cell death? Exp Cell Res 2003; 283:1–16.

Radiotherapy for early breast cancer. Cochrane Database Syst Rev 2002b CD003647.

Ravagnan L, Gurbuxani S, Susin SA, Maisse C, Daugas E, Zamzami N, Mak T, Jaattela M, Penninger JM, Garrido C, Kroemer G. Heat-shock protein 70 antagonizes apoptosis-inducing factor. Nat Cell Biol 2001; 3:839–843.

Real PJ, Cao Y, Wang R, Nikolovska-Coleska Z, Sanz-Ortiz J, Wang S, Fernandez-Luna, JL. Breast cancer cells can evade apoptosis-mediated selective killing by a novel small molecule inhibitor of Bcl-2. Cancer Res 2004; 64:7947–7953.

Riedl SJ, Salvesen GS. The apoptosome: signalling platform of cell death. Nat Rev Mol Cell Biol 2007; 8:405–413.

Riggins RB, Thomas KS, Ta HQ, Wen J, Davis RJ, Schuh NR, Donelan SS, Owen KA, Gibson MA, Shupnik MA, Silva CM, Parsons SJ, Clarke R, Bouton AH. Physical and functional interactions between Cas and c-Src induce tamoxifen resistance of breast cancer cells through pathways involving epidermal growth factor receptor and signal transducer and activator of transcription 5b. Cancer Res 2006; 66:7007–7015.

Riggins RB, Zwart A, Nehra R, Clarke R. The nuclear factor kappa B inhibitor parthenolide restores ICI 182,780 (Faslodex; fulvestrant)-induced apoptosis in antiestrogen-resistant breast cancer cells. Mol Cancer Ther 2005; 4:33–41.

Ron D. Translational control in the endoplasmic reticulum stress response. J Clin Invest 2002; 110:1383–1388.

Roninson IB, Broude EV, Chang BD. If not apoptosis, then what? Treatment-induced senescence and mitotic catastrophe in tumor cells. Drug Resist Updat 2001; 4:303–313.

Sachdev D, Li SL, Hartell JS, Fujita-Yamaguchi Y, Miller JS, Yee D. A chimeric humanized single-chain antibody against the type I insulin-like growth factor (IGF) receptor renders breast cancer cells refractory to the mitogenic effects of IGF-I. Cancer Res 2003; 63:627–635.

Sachdev D, Yee D. Disrupting insulin-like growth factor signaling as a potential cancer therapy. Mol Cancer Ther 2007; 6:1–12.

Saelens X, Festjens N, Vande WL, van GM, Van LG, Vandenabeele P. Toxic proteins released from mitochondria in cell death. Oncogene 2004; 23:2861–2874.

Salvesen GS. Caspases and apoptosis. Essays Biochem 2002; 38: 9–19.

Schiff PB, Fant J, Horwitz SB. Promotion of microtubule assembly in vitro by taxol. Nature 1979; 277:665–667.

Schneider BP, Sledge GW Jr. Drug insight: VEGF as a therapeutic target for breast cancer. Nat Clin Pract Oncol 2007; 4:181–189.

Shajahan AN, Wang A, Decker M, Minshall RD, Liu MC, Clarke R. Caveolin-1 tyrosine phosphorylation enhances paclitaxel-mediated cytotoxicity. J Biol Chem 2007; 282:5934–5943.

Siddik ZH. Biochemical and molecular mechanisms of cisplatin resistance. Cancer Treat Res 2002; 112:263–284.

Singal PK, Iliskovic N. Doxorubicin-induced cardiomyopathy. N Engl J Med 1998; 339:900–905.

Sledge GW Jr., Loehrer PJ Sr., Roth BJ, Einhorn LH. Cisplatin as first-line therapy for metastatic breast cancer. J Clin Oncol 1988; 6:1811–1814.

Speirs V, Carder PJ, Lane S, Dodwell D, Lansdown MR, Hanby AM. Oestrogen receptor beta: what it means for patients with breast cancer. Lancet Oncol 2004; 5:174–181.

Speirs V, Parkes AT, Kerin MJ, Walton DS, Carleton PJ, Fox JN, Atkin SL. Coexpression of estrogen receptor alpha and beta: poor prognostic factors in human breast cancer? Cancer Res 1999; 59:525–528.

Sriburi R, Jackowski S, Mori K, Brewer JW. XBP1: a link between the unfolded protein response, lipid biosynthesis, and biogenesis of the endoplasmic reticulum. J Cell Biol 2004; 167:35–41.

Tam WF, Sen R. IkappaB family members function by different mechanisms. J Biol Chem 2001; 276:7701–7704.

Thurlimann B, Keshaviah A, Coates AS, Mouridsen H, Mauriac L, Forbes JF, Paridaens R, Castiglione-Gertsch M, Gelber RD, Rabaglio M, Smith I, Wardley A, Price KN, Goldhirsch A. A comparison of letrozole and tamoxifen in postmenopausal women with early breast cancer. N Engl J Med 2005; 353:2747–2757.

Tolcher AW, Kuhn J, Schwartz G, Patnaik A, Hammond LA, Thompson I, Fingert H, Bushnell D, Malik S, Kreisberg J, Izbicka E, Smetzer L, Rowinsky EK. A Phase I pharmacokinetic and biological correlative study of oblimersen sodium (genasense, g3139), an antisense oligonucleotide to the bcl-2 mRNA, and of docetaxel in patients with hormone-refractory prostate cancer. Clin Cancer Res 2004; 10:5048–5057.

Tonko-Geymayer S, Doppler W. An essential link to mammary cancer? Nat Med 2002; 8:108–110.

Tu Y, Jerry DJ, Pazik B, Smith SS. Sensitivity to DNA damage is a common component of hormone-based strategies for protection of the mammary gland. Mol Cancer Res 2005; 3:435–442.

Ueda Y, Wang S, Dumont N, Yi JY, Koh Y, Arteaga CL. Overexpression of HER2 (erbB2) in human breast epithelial cells unmasks transforming growth factor beta-induced cell motility. J Biol Chem 2004; 279:24505–24513.

Vakkila J, Lotze MT. Inflammation and necrosis promote tumour growth. Nat Rev Immunol 2004; 4:641–648.

Vandenabeele P, Vanden BT, Festjens N. Caspase inhibitors promote alternative cell death pathways. Sci STKE 2006; 2006:pe44.

Vaux DL, Silke J. Mammalian mitochondrial IAP binding proteins. Biochem Biophys Res Commun 2003; 304: 499–504.

Voorhees PM, Orlowski RZ. The proteasome and proteasome inhibitors in cancer therapy. Annu Rev Pharmacol Toxicol 2006; 46:189–213.

Wang LG, Liu XM, Kreis W, Budman DR. The effect of antimicrotubule agents on signal transduction pathways of apoptosis: a review. Cancer Chemother Pharmacol 1999; 44:355–361.

Wang S, Konorev EA, Kotamraju S, Joseph J, Kalivendi S, Kalyanaraman B. Doxorubicin induces apoptosis in normal and tumor cells via distinctly different mechanisms. intermediacy of H(2)O(2)- and p53-dependent pathways. J Biol Chem 2004; 279:25535–25543.

Wang TH, Wang HS, Ichijo H, Giannakakou P, Foster JS, Fojo T, Wimalasena J. Microtubule-interfering agents activate c-Jun N-terminal kinase/stress-activated protein kinase through both Ras and apoptosis signal-regulating kinase pathways. J Biol Chem 1998; 273:4928–4936.

Wedam SB, Low JA, Yang SX, Chow CK, Choyke P, Danforth D, Hewitt SM, Berman A, Steinberg SM, Liewehr DJ, Plehn J, Doshi A, Thomasson D, McCarthy N, Koeppen H, Sherman M, Zujewski J, Camphausen K, Chen H, Swain SM. Antiangiogenic and antitumor effects of bevacizumab in patients with inflammatory and locally advanced breast cancer. J Clin Oncol 2006; 24:769–777.

Wilking N, Isaksson E, von SE. Tamoxifen and secondary tumours. An update. Drug Saf 1997; 16:104–117.

Witters LM, Witkoski A, Planas-Silva MD, Berger M, Viallet J, Lipton A. Synergistic inhibition of breast cancer cell lines with a dual inhibitor of EGFR-HER-2/neu and a Bcl-2 inhibitor. Oncol Rep 2007; 17:465–469.

Wolf BB, Green DR. Suicidal tendencies: apoptotic cell death by caspase family proteinases. J Biol Chem 1999; 274: 20049–20052.

Yang CH, Gonzalez-Angulo AM, Reuben JM, Booser DJ, Pusztai L, Krishnamurthy S, Esseltine D, Stec J, Broglio KR, Islam R, Hortobagyi GN, Cristofanilli M. Bortezomib (VELCADE) in metastatic breast cancer: pharmacodynamics, biological effects, and prediction of clinical benefits. Ann Oncol 2006; 17:813–817.

Yarden Y, Baselga J, Miles D. Molecular approach to breast cancer treatment. Semin Oncol 2004; 31:6–13.

Yaron A, Hatzubai A, Davis M, Lavon I, Amit S, Manning AM, Andersen JS, Mann M, Mercurio F, Ben-Neriah Y. Identification of the receptor component of the IkappaBalpha-ubiquitin ligase. Nature 1998; 396:590–594.

Yi JY, Shin I, Arteaga CL. Type I transforming growth factor beta receptor binds to and activates phosphatidylinositol 3-kinase. J Biol Chem 2005; 280:10870–10876.

Yoshida H, Matsui T, Yamamoto A, Okada, T, Mori K. XBP1 mRNA is induced by ATF6 and spliced by IRE1 in response to ER stress to produce a highly active transcription factor. Cell 2001; 107:881–891.

Yousefi S, Perozzo R, Schmid I, Ziemiecki A, Schaffner T, Scapozza L, Brunner T, Simon HU. Calpain-mediated cleavage of Atg5 switches autophagy to apoptosis. Nat Cell Biol 2006; 8:1124–1132.

Yu C, Friday BB, Lai JP, Yang L, Sarkaria J, Kay NE, Carter CA, Roberts LR, Kaufmann SH, Adjei AA. Cytotoxic synergy between the multikinase inhibitor sorafenib and the proteasome inhibitor bortezomib in vitro: induction of apoptosis through Akt and c-Jun NH2-terminal kinase pathways. Mol Cancer Ther 2006; 5:2378–2387.

Zhang K, Kaufman RJ. The unfolded protein response: a stress signaling pathway critical for health and disease. Neurology 2006; 66:S102–S109.

Zhang W, Ramdas L, Shen W, Song SW, Hu L, Hamilton SR. Apoptotic response to 5-fluorouracil treatment is mediated by reduced polyamines, non-autocrine Fas ligand and induced tumor necrosis factor receptor 2. Cancer Biol Ther 2003; 2:572–578.

Zhou BP, Hung MC. Dysregulation of cellular signaling by HER2/neu in breast cancer. Semin Oncol 2003; 30:38–48.

Zhu Y, Singh B, Hewitt S, Liu A, Gomez B, Wang A, Clarke R. Expression patterns among interferon regulatory factor-1, human X-box binding protein-1, nuclear factor kappa B, nucleophosmin, estrogen receptor-alpha and progesterone receptor proteins in breast cancer tissue microarrays. Int J Oncol 2006; 28:67–76.

Zimmermann S, Martens UM. Telomeres and telomerase as targets for cancer therapy. Cell Mol Life Sci 2007; 64:906–921.

BMC Bioinformatics

Methodology article

Knowledge-guided multi-scale independent component analysis for biomarker identification

Li Chen[1], Jianhua Xuan*[1], Chen Wang[1], Ie-Ming Shih[2], Yue Wang[1], Zhen Zhang[2], Eric Hoffman[3] and Robert Clarke[4]

Address: [1]Department of Electrical and Computer Engineering, Virginia Polytechnic Institute and State University, Arlington, VA, USA, [2]Departments of Pathology, Gynecology and Oncology, The Johns Hopkins University School of Medicine, Baltimore, MD, USA, [3]Research Center for Genetic Medicine, Children's National Medical Center, Washington, DC, USA and [4]Departments of Oncology and Physiology & Biophysics, Georgetown University School of Medicine, Washington, DC, USA

Email: Li Chen - lchen06@vt.edu; Jianhua Xuan* - xuan@vt.edu; Chen Wang - topsoil@vt.edu; Ie-Ming Shih - ishih@jhmi.edu; Yue Wang - yuewang@vt.edu; Zhen Zhang - zzhang7@jhmi.edu; Eric Hoffman - ehoffman@cnmcresearch.org; Robert Clarke - clarker@georgetown.edu

* Corresponding author

Published: 6 October 2008

BMC Bioinformatics 2008, **9**:416 doi:10.1186/1471-2105-9-416

Received: 23 May 2008
Accepted: 6 October 2008

This article is available from: http://www.biomedcentral.com/1471-2105/9/416

Abstract

Background: Many statistical methods have been proposed to identify disease biomarkers from gene expression profiles. However, from gene expression profile data alone, statistical methods often fail to identify biologically meaningful biomarkers related to a specific disease under study. In this paper, we develop a novel strategy, namely knowledge-guided multi-scale independent component analysis (ICA), to first infer regulatory signals and then identify biologically relevant biomarkers from microarray data.

Results: Since gene expression levels reflect the joint effect of several underlying biological functions, disease-specific biomarkers may be involved in several distinct biological functions. To identify disease-specific biomarkers that provide unique mechanistic insights, a meta-data "knowledge gene pool" (KGP) is first constructed from multiple data sources to provide important information on the likely functions (such as gene ontology information) and regulatory events (such as promoter responsive elements) associated with potential genes of interest. The gene expression and biological meta data associated with the members of the KGP can then be used to guide subsequent analysis. ICA is then applied to multi-scale gene clusters to reveal regulatory modes reflecting the underlying biological mechanisms. Finally disease-specific biomarkers are extracted by their weighted connectivity scores associated with the extracted regulatory modes. A statistical significance test is used to evaluate the significance of transcription factor enrichment for the extracted gene set based on motif information. We applied the proposed method to yeast cell cycle microarray data and Rsf-1-induced ovarian cancer microarray data. The results show that our knowledge-guided ICA approach can extract biologically meaningful regulatory modes and outperform several baseline methods for biomarker identification.

Conclusion: We have proposed a novel method, namely knowledge-guided multi-scale ICA, to identify disease-specific biomarkers. The goal is to infer knowledge-relevant regulatory signals and then identify corresponding biomarkers through a multi-scale strategy. The approach has been successfully applied to two expression profiling experiments to demonstrate its improved performance in extracting biologically meaningful and disease-related biomarkers. More importantly, the proposed approach shows promising results to infer novel biomarkers for ovarian cancer and extend current knowledge.

Background

Under their broadest definition, biomarkers include any biological or chemical indicator of a specific underlying process. In genetics, biomarkers are defined as a set of genes that are associated with a disease or are associated with the susceptibility to develop a specific disease. Microarray technology makes it possible to measure simultaneously the expression levels of thousands of genes, and identifying meaningful and useful biomarkers from these large data sets is a common goal. Specifically, investigators attempt to detect genes differentially expressed across different types of tissue samples or the samples obtained under different experimental conditions. Traditional biomarker identification methods have mainly been applied to statistical analysis of microarray data alone; T-test [1] and significance analysis of microarray (SAM) [2] are frequently used to detect differentially expressed genes between two phenotypes. Several new statistical methods have been developed to analyze time-course microarray data. Storey *et al*. proposed an algorithm (EDGE) to fit the time-course microarray data with natural cubic splines, followed by a goodness-of-fit test to detect differentially expressed genes [3]. Conesa *et al*. also proposed a two-step regression approach to sequentially identify differentially expressed genes from time-course microarray data under different conditions [4]. However, these and many related approaches do not incorporate knowledge of gene function, with respect to the phenotypes of interest, into their statistical models.

Ideally, biomarkers should not only exhibit differential gene expressions between normal and disease samples, but more importantly, they should also reflect their biological role in the disease phenotype. Most significance analysis methods applied to population (static) or time-course microarray data have the limitation that genes are analyzed independently and the interactions among them are ignored. Clustering methods, such as k-means clustering [5] and self-organizing maps (SOMs) [6], were introduced to group the genes with similar expression patterns. A shortcoming of the clustering methods is that they do not allow genes to be shared by multiple clusters. However, a single gene can be involved in multiple distinct biological processes [7]. One solution to this problem is to first infer gene regulatory networks [8-12] that appear to control or regulate phenotypically relevant biological functions, and then to extract the most biologically and statistically relevant biomarkers.

The application of Independent Component Analysis (ICA) to microarray data has shown some utility in regulatory network inference [10,13]. ICA is a statistically-principled linear decomposition method that models the observations as a linear combination of some latent (or hidden) variables [14]. From the perspective of a gene regulatory mechanism, any gene expression value can be regarded as a combinational effect of some regulatory inputs such as transcription factors, cellular functions, or responses to experiment conditions [10,12]. As demonstrated in our previous work [15,16] along with that of others [10,12], novel applications of ICA to high-throughput data from microarray technology can help reveal dominant regulatory mechanisms.

It is not a trivial task to link the estimated latent variables from ICA to real biological functions. To identify biologically relevant biomarkers for a specific disease, the incorporation of prior knowledge is of great importance to improving the accuracy of computational methods [17]. However, complete prior knowledge is often difficult to obtain. Some prior knowledge, such as regulatory motif information (promoter responsive element sequence) is available and can be incorporated into microarray data analysis to assist in regulatory module identification [18,19]. Recently, we have developed a new approach called motif-directed network component analysis (mNCA) to infer transcription regulatory activities (TFAs). This approach incorporates a stability analysis procedure to overcome the problem of many false positives in motif information [20]. Since we can only use known motifs, a clear limitation of the mNCA method is that we cannot infer any new potential regulatory biomarkers beyond prior knowledge from the model.

In this paper, we propose a novel method, namely knowledge-guided multi-scale ICA, to identify disease-specific biomarkers beyond partial prior knowledge. We propose that a latent variable estimated by ICA from the entire gene expression population represents the joint effect of several biological functions. Disease-specific biomarkers could be involved in several different biological functions by the ICA latent variables or linear regulatory modes. Therefore, we first cluster the whole gene population into multiple sub-populations in which only a few biological processes are involved. We then uncover the knowledge-relevant regulatory modes in each subpopulation based on the partial prior knowledge. Finally, disease-specific biomarkers are extracted according to the strength of their association with the extracted regulatory modes. A statistical test is applied to evaluate the significant enrichment of transcription factors for the extracted biomarkers based on motif information.

For algorithm validation, we applied our approach to two time-course microarray data sets to demonstrate its improved performance. The first data set is a yeast cell cycle microarray data set with 104 well known cell cycle-related genes; the second is a remodeling and spacing factor 1 (Rsf-1) induced microarray data set from a profiling study of ovarian cancer. The experimental results show

that our approach can identify biologically meaningful disease-specific biomarkers related to ovarian cancer, as compared to other gene selection methods with or without prior knowledge.

Methods

If we apply ICA directly onto an entire gene expression population, the extracted regulatory modes will reflect the joint effect of several biological functions, some of which are related to the disease under study and some are not. To overcome this problem, we developed a divide-and-conquer strategy. We applied a knowledge-guided multi-scale ICA approach to extract disease-related regulatory modes reliably, and then we identify the biomarkers associated with the modes. The overall scheme is illustrated in Fig. 1. Firstly, a knowledge gene pool (KGP) is constructed by collecting the genes that are known to be relevant to the specific disease from available databases and literatures. Secondly, the entire gene population is divided into sub-populations by a clustering method applied to the micro-array data and, to identify regulatory modes, ICA is applied to each sub-population. The most relevant linear regulatory mode in each cluster is extracted using the gene metadata in the KGP and the associated biomarkers are ranked according to their weighted loading factors. Finally, motif enrichment analysis is conducted to evaluate the extracted biomarker candidates in terms of the enrichment of disease-related transcription factors.

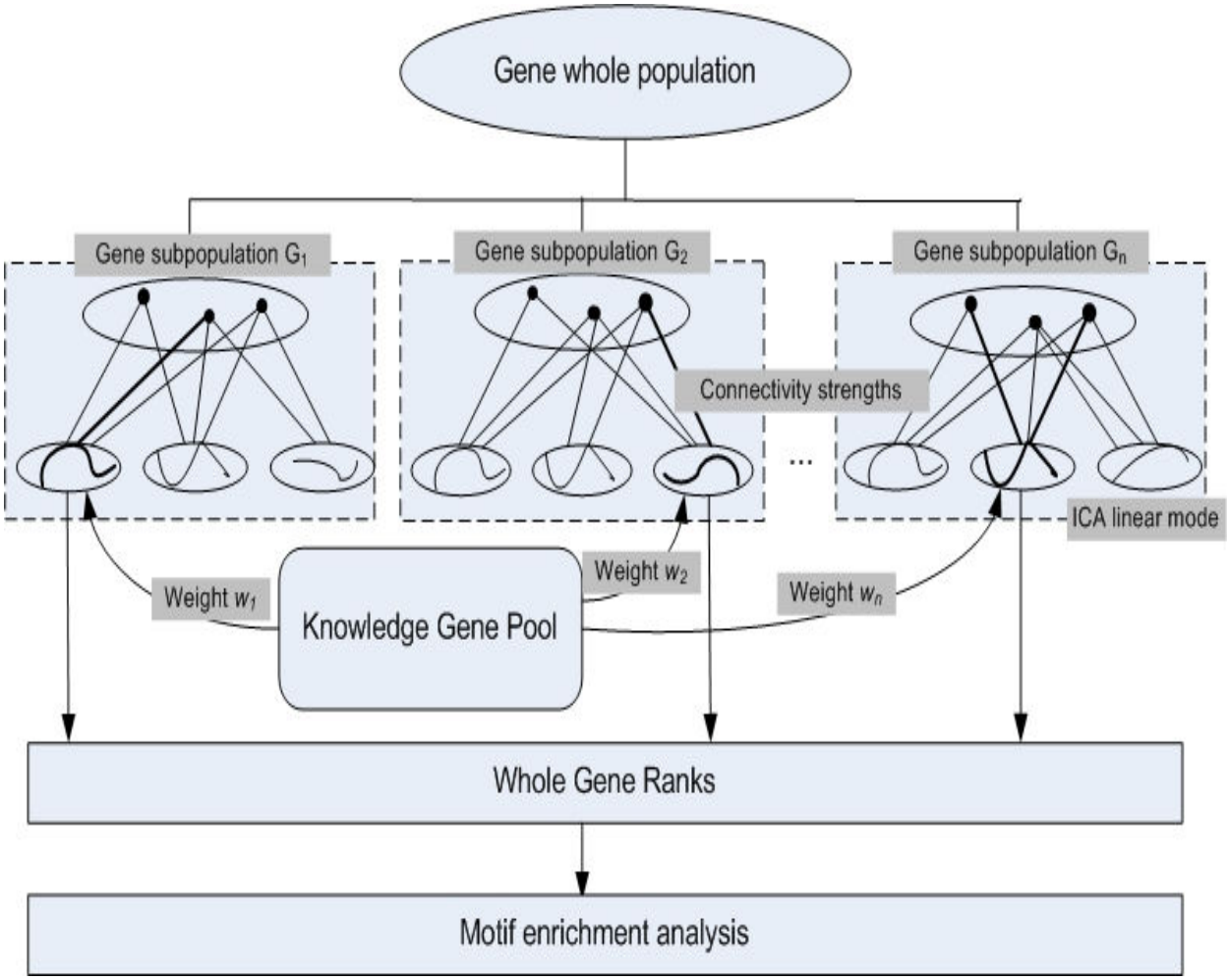


Figure 1
Flow chart of the proposed method – knowledge-guided multi-scale independent component analysis (ICA) – for biomarker identification.

Independent component analysis (ICA)

Consider a gene expression data matrix $\mathbf{X} = [x_{ji}]$, whose rows correspond to different microarray samples, and columns correspond to individual genes. ICA decomposition model can be mathematically formulated as (assuming noiselessness for simplicity):

$$\mathbf{X}_{N \times L} = A_{N \times M}\mathbf{S}_{M \times L}, \quad (1)$$

$$\mathbf{U}_{M \times L} = W_{M \times N}\mathbf{X}_{N \times L}, \quad (2)$$

where Equation (1) describes the linear combination model with mixing matrix $\boldsymbol{A}$, and Equation (2) the decomposition model with de-mixing matrix $\boldsymbol{W}$. $\boldsymbol{S}$, $\mathbf{X}$ and $\mathbf{U}$ are independent components, mixtures, and estimated independent components, respectively. M is the number of independent components, N the number of samples and L the number of genes.

In microarray data analysis, an ICA model could be interpreted as the expression value of an individual gene i under condition j ($\mathbf{x}_i(j)$) is the summation of different linear modes in $\boldsymbol{A}$ at condition j ($\mathbf{a}_k(j)$) weighted by independent loading factors s_{ik} in $\boldsymbol{S}$[8], as shown below:

$$\mathbf{x}_i(j) = \sum_{k=1}^{M} s_{ik}\mathbf{a}_k(j), \quad i = 1, ..., L; j = 1, ..., N. \quad (3)$$

The linear modes in $\boldsymbol{A}$ might reflect distinct regulatory mechanisms involved in gene regulation, such as transcription factor (TF) activities. The FastICA algorithm [21] can be utilized to obtain $\boldsymbol{A}$ and $\boldsymbol{S}$ based on the assumption that the components are statistically independent and have non-normal distributions (typically super-Gaussian). This assumption is biologically plausible as most genes are not expected to change dramatically. Only the genes involved in distinct regulatory mechanisms will change, producing super-Gaussian distributions in microarray data.

Several methods have been developed to associate a set of genes with a specific linear mode [10,12,22]. These methods each assume that genes with the highest absolute loading values are the significant genes associated with linear mode $\mathbf{a}_k$. In this paper, genes are ranked by a modified criterion based on the same assumption as described in the next subsection.

Knowledge-guided multi-scale ICA

Since ICA is an unsupervised method, it is difficult to determine which linear modes are related to specific biological functions. To identify the biomarkers relevant to a specific biological function, prior knowledge could provide guidance for any computational method. In this approach, we will collect a KGP containing genes strongly associated with the disease and use these to guide the ICA approach for disease-relevant biomarker identification. Notice that the total connection strength of the knowledge genes associated with a disease-relevant linear mode would be larger, in principle, than that of irrelevant linear modes. Based on this observation, the most knowledge-relevant linear mode can be determined from the estimated ICA modes and the associated genes can then be extracted.

However, if we apply ICA to the entire molecular profile, the estimated linear modes will likely reflect the joint effect of several biological functions, even for the most knowledge-relevant mode, because many disease-irrelevant but differentially expressed genes co-exist in the data. Conversely, biomarkers should be involved in several different linear modes in relation to underlying biological processes. Therefore, it is reasonable to first separate the entire profile into sub-populations. We can then find the specific ICA linear modes from different subsets of genes rather than from the whole gene population; this approach is referred to as the "multi-scale ICA" approach in this paper. Since these modes will be associated with different parts of the knowledge genes in the KGP, they are more suitable for biomarker identification. Clustering methods, such as k-means clusterin and SOMs, can be used to form the subsets of genes, with the assumption that the genes involved in similar biological functions are more likely to exhibit similar expression patterns than genes involved in different biological functions.

Our method can be mathematically described as follows. Assume a whole gene population $\boldsymbol{G}$ in a microarray data $\mathbf{X}$ has been clustered into n subsets, $\boldsymbol{G}_1$, $\boldsymbol{G}_2$, ..., $\boldsymbol{G}_n$. For each subset $\boldsymbol{G}_i$ (i = 1, ..., n), we apply ICA to find the most knowledge-relevant linear mode $\mathbf{a}_j$ according to the total connection strength of the knowledge genes in this subset. Thus, the index j can be obtained by

$$j = \arg\max_{m}\left(\sum_{g \in K_i} \left| s_{gm} \right|\right) \quad m = 1, ..., M_i, \quad (4)$$

where s_{gm} is the loading factor for gene g associated with linear mode $\mathbf{a}_m$, K_i the subset of knowledge genes in the i^{th} cluster, and M_i the number of independent components in the i^{th} cluster.

Then each gene g in this subset is assigned a score c, which is defined as follows:

$$c_g = w_i * \left| s_{gj} \right|, \quad g \in G_i, \quad w_i = \frac{\left| K_i \right|}{\left| K \right|}, \quad (5)$$

where w_i is a weight to represent the significance of the linear mode in the i^{th} subset associated with the prior knowl-

edge. Here we define w_i as the proportion of all knowledge genes in this subset with respect to the entire KGP (K). Once the knowledge-relevant linear modes in all subsets are determined, each gene will have a score assigned and we rank the genes in terms of their scores. The larger the score, the more strongly the gene is related to the biological process.

A key issue in this method is how to determine the optimal cluster number when forming the subsets of genes. In this paper, we determine the optimal cluster number by a cross-validation approach. Specifically, we assume the optimal cluster number is in some range, from 1 to an upper limit. For each cluster number, the knowledge genes are randomly stratified into a training gene set (as our partial prior knowledge gene set) and a test gene set by a ten-fold cross-validation approach. The method is applied with the partial prior knowledge genes to rank the whole gene population, and prediction accuracy is tested on the test gene set. The above procedure is repeated 10 times, once for each left out fold, and an average accuracy over the ten folds is reported. We select the number with the highest average accuracy as the optimal cluster number for clustering. The upper limit of cluster numbers should be cautiously determined by the number of knowledge genes and the number of genes in the full profile. If the number of clusters is too large, it will lose the ability to infer novel biomarkers. An extreme case is that each individual gene forms a cluster and then we can only obtain the correct ranks for known genes. Genes not in the KGP will be randomly ranked, which is not informative at all for biomarker identification. If the cluster number is too small, the estimated linear modes may be incorrect due to the presence of many irrelevant genes. In our experiments, we set the upper limit as 10 for the yeast cell cycle data set and 15 for the ovarian cancer microarray data set, respectively.

Knowledge gene pool (KGP)

Each KGP is a collection of those genes that are potentially most strongly related to a specific disease. Usually there are thousands of genes in microarray data and most of them are not relevant to a specific disease even though they exhibit changes in gene expression level. The knowledge gene pool is an important asset for data analysis since it helps reduce many false positives. However, in most cases, little prior knowledge can be obtained, and the available knowledge is usually neither complete nor sufficiently accurate to fully define the specific disease under study. Thus, the KGP is best used as a guide for biomarker identification. In our studies, the KGP is primarily constructed from the published biological literature or from databases such as Ingenuity Pathway Analysis (IPA; Ingenuity Systems: http://www.ingenuity.com) and the TRANSFAC 11.1 Professional Database [23].

Evaluation by motif enrichment analysis

For microarray data analysis, there is often no ground truth (i.e., true biomarkers known to be related to a specific biological process or disease under study) available for us to evaluate the performance of a biomarker identification method. However, we know that gene expression is often regulated by transcription factors (TFs), proteins that bind to promoter or enhancer sequence elements upstream of genes and either activate or inhibit gene expression. In this paper, with the motif information provided, we have designed a statistical test to evaluate the enrichment of transcription factors for a gene set identified. A gene-transcription factor matrix $\boldsymbol{M}$ is generated where each element in the matrix, m_{gf}, represents how well the upstream sequence of a gene g matches the motif that a transcription factor f binds to. For human genes, 2 Kbp upstream regions from the transcription start sites (TSSs) of the genes are extracted from the UCSC genome databases [24]. Match™ [25] is then used to search the transcription factor binding site (TFBS) by its position-weighted matrices (PWMs) in a gene's upstream region, which outputs the scores of core similarity and matrix similarity for each matched motif. Since one TF may have multiple TFBSs, we use the summation of average scores of core similarity and matrix similarity to set the final value of m_{gf}.

Given a gene set S extracted by a computational method, a statistic to measure the enrichment of a specific transcription factor f is defined as

$$e_f = \sum_{g \in S} m_{gf}. \tag{6}$$

To calculate the statistical significance (p-value), we need to form a null distribution. The null hypothesis is that the gene set is randomly generated from the gene population and there is no significant enrichment of the transcription factor f. We randomly select gene sets with same size of S from the baseline gene population, and repeat B times to generate the corresponding null statistic enrichment score e_f^{0b}, for b = 1, ..., B. The null hypothesis distribution is assumed to be symmetric in this study. The p-value can be obtained for each gene set by calculating the probability that a null gene set has a statistic more extreme than the observed statistic. Mathematically, the p-value can be calculated by:

$$p_S = \frac{\text{number of members in } \{b: e_f^{0b} > e_f, b=1,\ldots,B\}}{B}. \tag{7}$$

Baseline experiments and evaluation method

To evaluate the performance of our proposed approach, EDGE algorithm [3] was first considered as a comparison method since it was specially designed to identify statistically significant genes from time-course microarray data. However this comparison is insufficient due to that EDGE does not incorporate knowledge genes to provide guidance for biomarker identification. On the other hand, given partial prior knowledge genes, traditional supervised classification methods are not suitable to predict whether a gene is related to prior knowledge because there is no true negative gene available. Therefore, we design three baseline biomarker identification methods that incorporate partial prior knowledge for a fair comparison. The first baseline ICA method is designed to evaluate if our multi-scale strategy by clustering offers an improved performance for biomarker identification. Two correlation methods with or without clustering are then implemented to identify the genes exhibiting similar patterns with partial prior genes, compared to the ICA approach focusing on regulatory mode identification. Specifically, the first method is a baseline ICA method where ICA is applied to the entire expression profile and the partial prior knowledge is used to find the most knowledge-relevant linear mode by Equation (4). Genes are ranked according to their absolute connection strengths associated with this linear mode. The second method estimates the correlation with the partial prior knowledge genes without clustering (baseline correlation method-1). Genes are then ranked based on their absolute correlation coefficients between an individual gene expression profile and the average profile of partial prior knowledge genes. However, taking the average profile of all knowledge genes may reduce the sensitivity of detection, especially when the genes in KGP are not similar to each other. To overcome this problem, the third baseline method is a weighted correlation method based on a clustering approach (baseline correlation method-2). Similar to the multi-scale ICA method, the entire gene population is grouped into several sub-populations and a gene in each cluster is assigned a score. The score is the weighted absolute correlation coefficient between an individual gene expression profile and the average profile of partial prior knowledge genes in this cluster. The weight is then calculated using Equation (5) and genes are ranked according to their scores.

Given a ranked gene list and knowledge gene set, we can use the Receiver Operating Characteristic (ROC) curve [26] and the area under the curve (AUC) to measure the test accuracy for each biomarker identification method. ROC curve is a graphical plot of true positive rate (TPR) vs. false positive rate (FPR). AUC is an important performance measure that provides an overall measure of accuracy for the test. Given a ranked gene list (g_1, g_2, ..., g_n) with a total of n genes and the ground truth gene set G_k with k genes, true positive rate and false positive rate, when selecting top i genes G_i in the list, are calculated as follows:

$$TPR(i) = \frac{|G_i \cap G_k|}{k}, \quad (8)$$

$$FPR(i) = \frac{i - |G_i \cap G_k|}{n - k}. \quad (9)$$

Results and discussion

We applied our knowledge-guided multi-scale ICA method to two gene expression profiling studies: (1) a yeast cell cycle microarray data set [27] and (2) an Rsf-1-induced microarray data set. The yeast cell cycle data set consists of the expression of 6178 Open Reading Frames (ORFs) during the cell replication cycle in the budding yeast (Saccharomyces cerevisiae). The data set consists of 77 samples corresponding to various experiment conditions. Approximately 800 genes have been identified as cycle-regulated genes; among these 104 genes have been well studied [27]. We us The goal of this experiment is to identify the cell cycle-regulated linear modes and then extract the corresponding genes associated with the cell cycle. We used the 104 genes as our training knowledge gene set and the remaining 704 genes as an independent test set for evaluation.

The Rsf-1-induced microarray data set was acquired and analyzed in our experiment. The dataset was generated using Affymetrix Human Genome U133 Plus 2.0 Arrays from an expression profiling study of ovarian cancer at the Johns Hopkins Medical Institutions. The study was designed to identify Rsf-1 regulated genes in ovarian cancer; Rsf-1 (also known as HBXAP) is a newly discovered gene frequently amplified in ovarian cancer [28]; the protein participates in chromatin remodeling which is essential for a variety of cellular functions including transcription, DNA replication, and DNA repair. The data set is composed of 7 samples with two biological conditions (Rsf-1-induced and not Rsf-1-induced) and four time points at 0 hour, 6 hours, 18 hours, and 30 hours. We used Affymetrix's Probe Logarithmic Intensity Error (PLIER) algorithm with quantile normalization to preprocess the original intensity data for gene expression measurements [29]. After the preprocessing, we obtained expression measurements of 54,675 probe sets for each sample.

The EDGE algorithm was first applied to select statistically significant expressed genes from yeast cell cycle data and Rsf-1 induced ovarian cancer data, respectively. After ranking all genes in terms of their q-values estimated from EDGE, we calculated AUC values for yeast cell cycle-related genes and ovarian cancer-related genes, respec-

tively (see below). As a result, both AUC values are relatively low (around 0.5), which indicates that the genes identified from pure data-driven methods (such as EDGE; without prior knowledge guidance) may not show strong biological relevance.

Yeast cell cycle data

To reduce computational complexity, k-means clustering was used to form the subsets of genes for both datasets. The number of independent components in the FastICA algorithm was set to five for this dataset, since our previous dimension estimation approach with a stability analysis procedure [16] showed that five independent components are sufficient to describe the gene expression data. We first conducted ten-fold cross-validation on the well studied 104 cell cycle-related genes. For each fold, the optimal cluster number is determined by a nested cross-validation procedure on the training gene set, as illustrated in Fig. 2. The number of clusters ranges from 1 to 10. Notice that when the number is 1, no clustering is needed and the algorithm reduces to the baseline ICA method. Each ten-fold cross-validation is repeated 10-times with different randomly chosen stratified sets of knowledge genes. Since the k-means clustering method generates different results depending on its random initialization, we repeat the procedure ten times with different initializations to obtain more reliable results. The results reported here are the average results of the ten different initializations.

The resulting average AUC value of ten-fold cross-validation on 104 genes is 0.9206 with standard deviation of 0.0470. Fig. 3 shows the histogram of determined optimal number of clusters during the ten-fold cross-validation procedure. From the figure we can see the most frequent number of clusters is five. Then we implemented three baseline methods for ten-fold cross-validation as comparisons. For baseline correlation method-2, we chose the optimal cluster number from the multi-scale ICA method for a fair comparison. The ROC curves of ten-fold cross validation for the two baseline correlation methods, the baseline ICA method, and our multi-scale ICA method are shown in Fig. 4. The ROC curves show that the multi-scale ICA method outperforms the baseline correlation method-2, and that the baseline ICA approach is better than the baseline correlation method-1. Overall, the proposed multi-scale ICA method significantly outperforms all three baseline methods as estimated by the Kolmogorov-Smirnov (K-S) one-sided test (Table 1).

To further test the generalizability of our method, we conducted ten-fold cross-validation on the 104 genes using a subset of samples. The original data set includes 77 samples synchronized by three independent methods: α factor arrest, elutriation and arrest of a cdc 15 temperature-sensitive mutant [27]. We selected 63 samples from all the

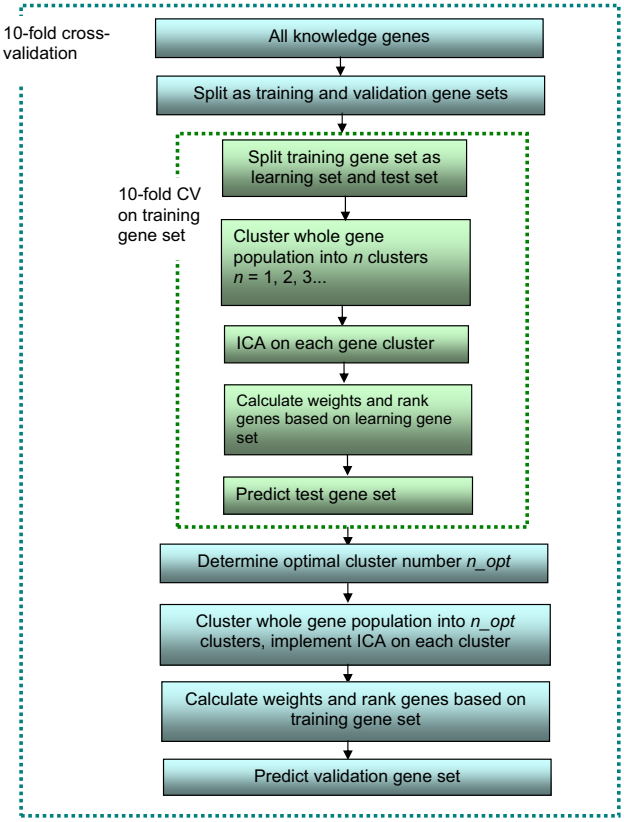


Figure 2
Procedure of ten-fold cross-validation. The optimal number of clusters is determined by a nested ten-fold cross-validation on training gene set.

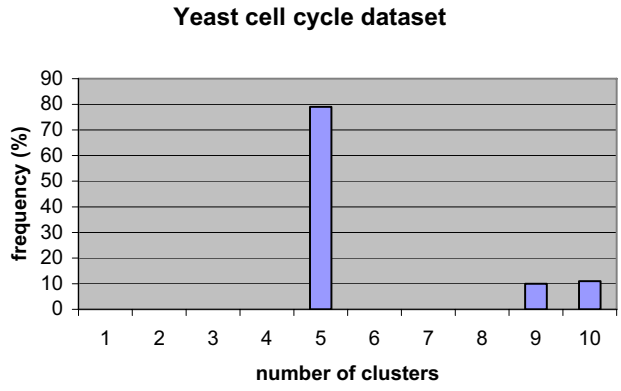


Figure 3
Histogram of determined optimal number of clusters in ten-fold cross- validation on yeast cell cycle data set.

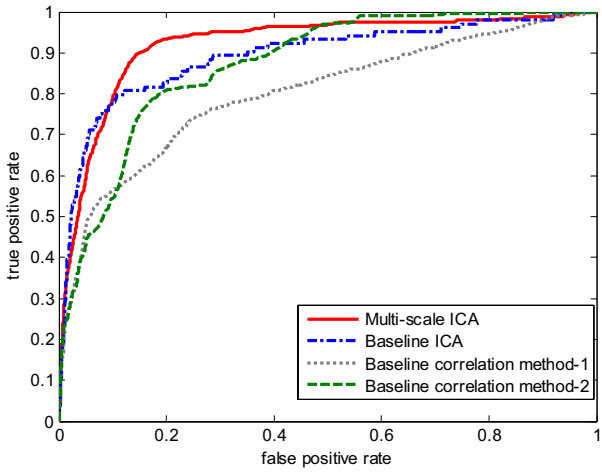


Figure 4
ROC curves of ten-fold cross-validation for four biomarker identification methods on training knowledge gene set of yeast cell cycle data set. Solid line represents the multi-scale ICA method; dash-dotted line represents the baseline ICA method; dotted line represents the correlation method-1; dash line represents the correlation method-2.

samples by excluding those samples under elutriation condition. The resulting average AUC value is 0.9157 with standard deviation of 0.0458. Also the most frequent optimal cluster number is five (with a frequency of 65%), which shows a great consistency when compared to the result using all the samples.

All 104 knowledge genes were then used as a training set in the algorithm to test 704 cell cycle-related genes for all four methods. During the training, we still used tenfold cross-validation to determine the optimal number of clusters. Fig. 5 shows the average AUC values and their standard deviations in ten-fold cross-validation across different number of clusters. From the figure we can see that the average AUC (standard deviation), starting at 0.892 (0.0006) for the full gene population, decreases a little at two and three clusters. The AUC increases gradually and reaches the peak of 0.9274 (0.0071) at five clusters, at which it remains constant. So the optimal number of clusters for multi-scale ICA approach is five. Then an independent evaluation was performed on the test gene set and the ROC curves for these four methods was calculated when the cluster number is five (Fig. 6). The ICA-based methods significantly outperform the baseline correlation methods, and the multi-scale ICA is the best method when compared with the three baseline methods (Table 2).

Table 1: P-values of Kolmogorov-Smirnov test for different methods on yeast cell cycle data using ten-fold cross-validation

Method 1	Method 2	P-values of K-S test
Optimal ICA	Baseline ICA	< 1e-10
Optimal ICA	Correlation method 1	< 1e-10
Optimal ICA	Correlation method 2	< 1e-5

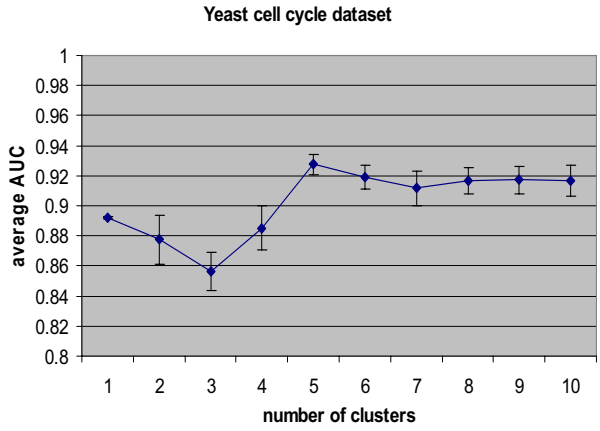


Figure 5
Average area under the curve (AUC) values using ten-fold cross-validation with different numbers of clusters on 104 knowledge genes. The knowledge-guided multi-scale ICA method is applied to yeast cell cycle data set for the identification of cell cycle-related genes.

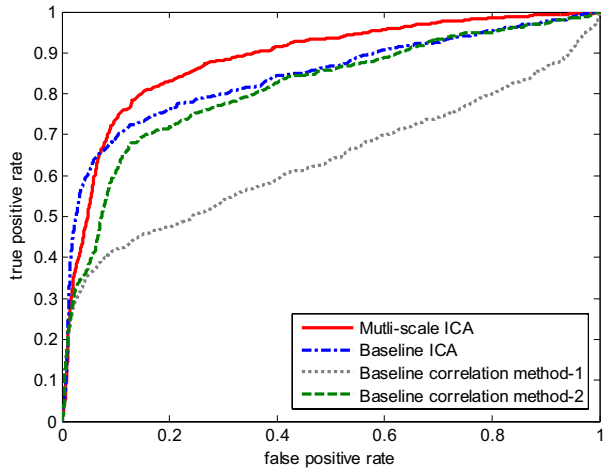


Figure 6
ROC curves of four biomarker identification methods on yeast cell cycle data set with an independent test gene set.

Table 2: P-values of kolmogorov-Smirnov test for different methods on yeast cell cycle data using an independent test gene set

Method 1	Method 2	P-value of K-S test
Optimal ICA	Baseline ICA	< 1e-10
Optimal ICA	Correlation method 1	< 1e-10
Optimal ICA	Correlation method 2	< 1e-10

We examined in detail the extracted knowledge-relevant linear modes and the biological functions of their associated cell cycle-regulated genes. Fig. 7 shows five knowledge-relevant linear modes and their weights as identified when the number of clusters is set at the optimum number of five (Fig. 5). The top three linear modes have much higher weights than the lower two modes and their estimated TFAs clearly show periodic patterns related to cell cycle. We examined the biological functions of these well-known cell cycle-regulated genes associated with these three linear modes. The majorities of genes in linear mode L3 are associated with the M/G1 boundary or are known transcriptional targets of STE12/MCM1. Most of

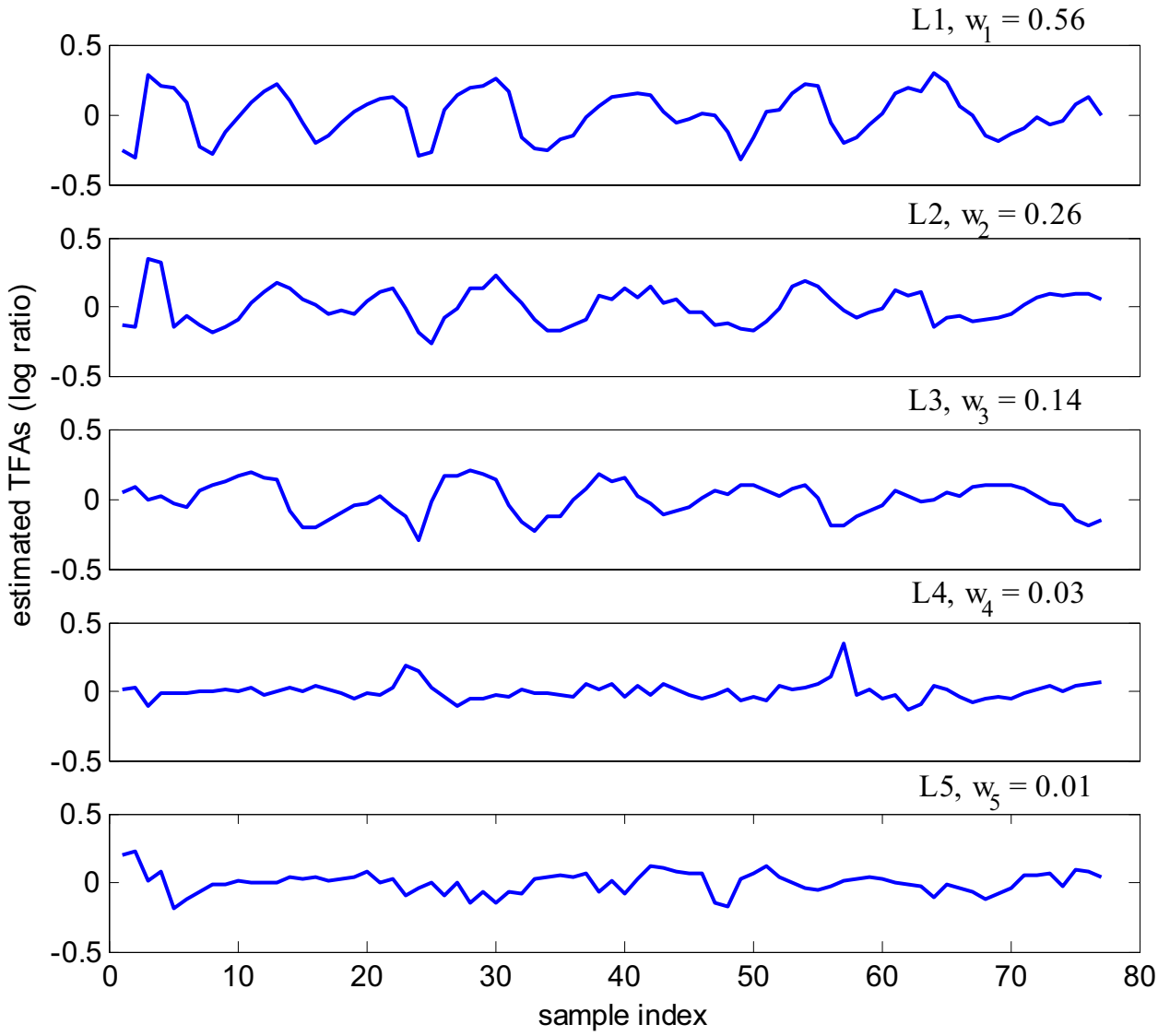


Figure 7
Five cell cycle-related linear modes in the proposed multi-scale ICA approach on yeast cell cycle data set. The weight is also listed in the figure for each linear mode.

the genes in linear mode L1 are SCB/MCB regulated in late G1 and S phase. Finally, many genes in linear mode L2 are in S/G2 and G2/M phases. In summary, we can see that the linear modes L3, L1, and L2 correspond to different biological functions in cell cycle process.

The top 10 genes selected by multi-scale ICA method are listed in Table 3. Among them, four genes (CLN2, MCD1, POL30 and RNR1) are in the known training gene set. All other genes (CSI2, PRY2, YOX1, TOS4, AXL2 and CRH1) are the genes related to cell cycle beyond our training gene set, i.e., in the test gene set. The results show that our method is effective at finding novel biomarkers beyond knowledge, which is clearly an important feature of the proposed approach for novel biomarker identification beyond prior knowledge. In most of cases, especially for human disease, knowledge genes are limited and we need to infer the new ones from partial knowledge for biomarker discovery.

Rsf-1-induced gene expression data

Knowledge gene pool (KGP)

To construct the KGP, we started with the known gene Rsf-1 and its related genes, NF-kappa B (NFKB1) and SMARCA5 (also known as hSNF2H) as reported in [30], to search the databases. We used Ingenuity Pathway Analysis (IPA) to extract 95 genes that are thought to be directly related to NFKB1 and SMARCA5. Note that there is no network related to Rsf-1 in the current IPA database. We also included 43 genes from TRANSFAC 11.1 Professional Database [23], whose protein products are transcription factors biologically relevant to ovarian cancer as reported in literature. Hence, our KGP consists of 141 distinct Affymetrix probe set identifiers that represent the expression values for the 138 genes.

Multi-scale ICA results

We used 'tanh' nonlinearity in the FastICA algorithm: other parameters were set at their default values. The number of the independent components is set to a maximum value of 6 due to the limitation of sample size. Ten-fold cross-validation was conducted on our partial prior knowledge genes, where the optimal cluster number was determined by a nested cross-validation approached for each fold as shown in Fig. 2. The number of clusters was set from 1 to 15. We also repeated 10 times for ten-fold cross-validation and k-means clustering in order to generate more reliable results. The resulting average AUC is 0.7203 with standard deviation of 0.0804. Fig. 8 shows the histogram of determined optimal cluster number in the ten-fold cross-validation procedure and we can see that the most frequent cluster number is 4. We compared the ROC curves for the two baseline correlation methods, the baseline ICA and the multi-scale ICA for ten-fold cross-validation (Fig. 9). The results in Table 4 show that multi-scale ICA method performs significantly better than baseline ICA method and baseline correlation method-1 with p-value < 1e-10, while performing marginally better than baseline correlation method-2 (p-value = 0.0037). Since baseline correlation method-2 also calculates clustered average profiles of the prior knowledge genes, this result indicates that the multi-scale approach by clustering is an effective strategy to improve the performance for ovarian cancer-related biomarker identification. On the other hand, a major weakness in baseline correlation method-1 lies in that the average profile of all prior knowledge is used when their expression profiles are not similar to each other.

Evaluation by motif analysis

All knowledge genes were used as the training set in the algorithm to rank the whole gene population for all four methods. During the training, we still used ten-fold cross-validation to determine the optimal number of clusters in multi-scale ICA method. Fig. 10 shows the average AUC values and their standard deviations obtained with different numbers of clusters for the ten-fold cross-validation; the average AUC (standard deviation), starting at 0.6146 (0.0004) for the whole gene population, increases to 0.7329 (0.0253) at two clusters and reaches the maximum

Table 3: Top10 genes selected by the proposed multi-scale ICA method on yeast cell cycle data

Rank	ORF	Name	Short Description
1	YPL256C	CLN2	CycLiN; G1 cyclin involved in regulation of the cell cycle
2	YOL007C	CSI2	Chitin Synthesis Involved; protein of unknown function
3	YKR013W	PRY2	Pathogen Related in Yeast; protein of unknown function
4	YDL003W	MCD1	Mitotic Chromosome Determinant; expression is cell cycle regulated and peaks in S phase
5	YML027W	YOX1	Homeodomain-containing transcriptional repressor
6	YBR088C	POL30	POLymerase; proliferating cell nuclear antigen (PCNA)
7	YLR183C	TOS4	Target of SBF; promoters of some genes involved in pheromone response and cell cycle;
8	YIL140W	AXL2	AXiaL budding pattern; glycosylated by Pmt4p; potential Cdc28p substrate
9	YGR189C	CRH1	Congo Red Hypersensitive; cell wall protein; putative chitin transglycosidase
10	YER070W	RNR1	RiboNucleotide Reductase; the RNR complex catalyzes the rate-limiting step in dNTP synthesis and is regulated by DNA replication and DNA damage checkpoint pathways via localization of the small subunits

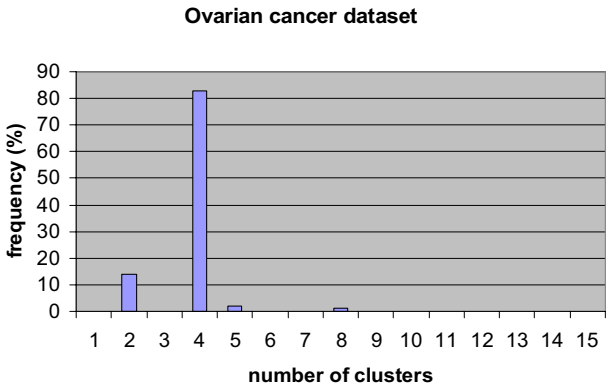


Figure 8
Histogram of determined optimal number of clusters in ten-fold cross- validation on ovarian cancer data set.

value of 0.7343 (0.0210) at four clusters, and remains almost constant thereafter. Therefore, the optimal number of cluster for the multi-scale ICA approach was selected as four. Specifically, we examined estimated linear modes from ICA methods. Fig. 11 shows the estimated knowledge-related TFAs using baseline ICA method and Fig. 12 shows the estimated four knowledge-related TFAs and their weights using our multi-scale ICA method. We observe that one of the TFA patterns in Fig. 12 (L3) is similar with that in Fig. 11, which indicates that multi-scale ICA method can estimate more TFAs for knowledge-related genes than baseline ICA method. Four different linear modes and their weights in Fig. 12 also indicate that the expression patterns of the genes in KGP are not similar to each other, which seems to be the major reason behind that baseline correlation method-1 (using the average profile of all prior knowledge) underperforms other methods.

For the final ranked gene lists, we performed motif enrichment analysis to evaluate the performance of each of the four different methods for biomarker identification. Specifically, among 43 ovarian cancer-related TFs extracted from TRANSFAC 11.1 Professional Database [23], 14 TFs have their PWMs available and we generated the gene-TF matrix M for them. For each TF, a PWM was chosen from

Table 4: P-values of Kolmogorov-Smirnov test for different methods on Rsf-1-induced ovarian cancer microarray data

Method 1	Method 2	p-value of the K-S test
Optimal ICA	Baseline ICA	< 1e-10
Optimal ICA	Correlation method 1	< 1e-10
Optimal ICA	Correlation method 2	0.0037

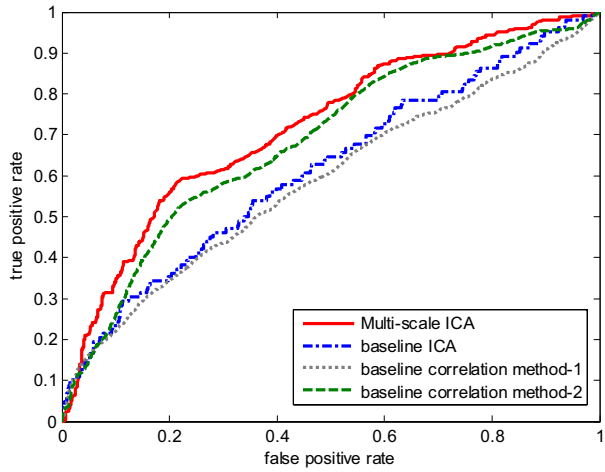


Figure 9
ROC curves of ten-fold cross-validation for four biomarker identification methods on knowledge gene set of ovarian cancer data set. Solid line represents the multi-scale ICA method; dash-dotted line represents the baseline ICA method; dotted line represents the correlation method-1; dash line represents the correlation method-2.

the vertebrate non-redundant profiles. Table 5 lists their TRANSFAC PWM entry IDs and the corresponding TF descriptions. To increase the statistical power, we conducted multiple tests by selecting different gene sets with different sizes for different gene selection methods. The number of genes in each gene set ranges from 100 to 1,000 and the average p-values for 14 TFs are reported.

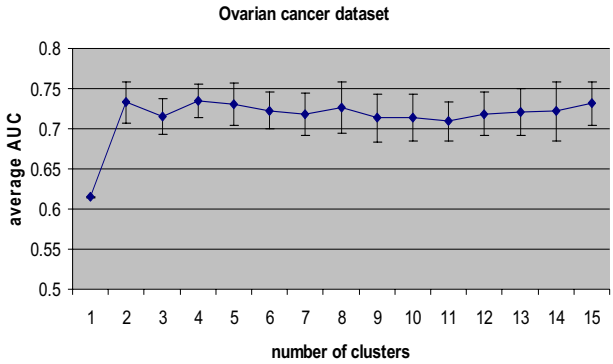


Figure 10
Average AUC values using ten-fold cross-validation across different numbers of clusters. The knowledge-guided multi-scale ICA method is applied to Rsf-1-induced ovarian cancer microarray data set for the identification of disease-specific biomarkers.

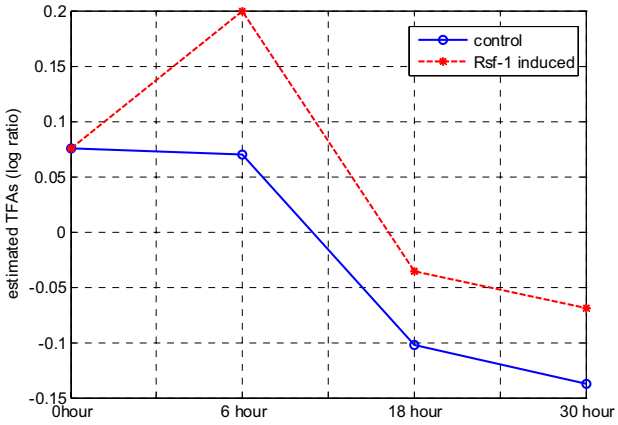


Figure 11
Estimated knowledge-related TFAs using baseline ICA method. X-axis represents the time and Y-axis represents the estimated TFAs.

Fig. 13 shows the average p-values of TFs enrichment for different gene sets selected by different methods. Both ICA methods outperform the baseline correlation methods in terms of finding more enriched ovarian cancer-related TFs binding sites. Moreover, our multi-scale ICA method is slightly better than baseline ICA method for motif enrichment. It is worth noting that although both multi-scale ICA and baseline ICA methods can extract ovarian cancer-related biomarkers with significant motif enrichment, multi-scale ICA method can help reveal more biomarkers related to ovarian cancer. For this experiment, it is also expected to have similar TF enrichment from both methods, since one common linear mode is revealed by both methods (i.e., the mode in Fig. 11 is very similar with the L3 mode in Fig. 12). From the pattern of this common mode, we postulate that this is a major mode related to RSF-1-induced ovarian cancer. Therefore, the genes extracted from this mode will show a similar significance level in TF enrichment (as shown in Fig. 13). However, the multi-level ICA approach can extract other linear modes related to ovarian cancer (see Fig. 12). Apparently, the biomarkers related to these other modes cannot be identified with the baseline ICA approach. This can be supported by the ROC curves in Fig. 9, showing an improved performance of using multi-scale ICA approach compared to that of using baseline ICA approach.

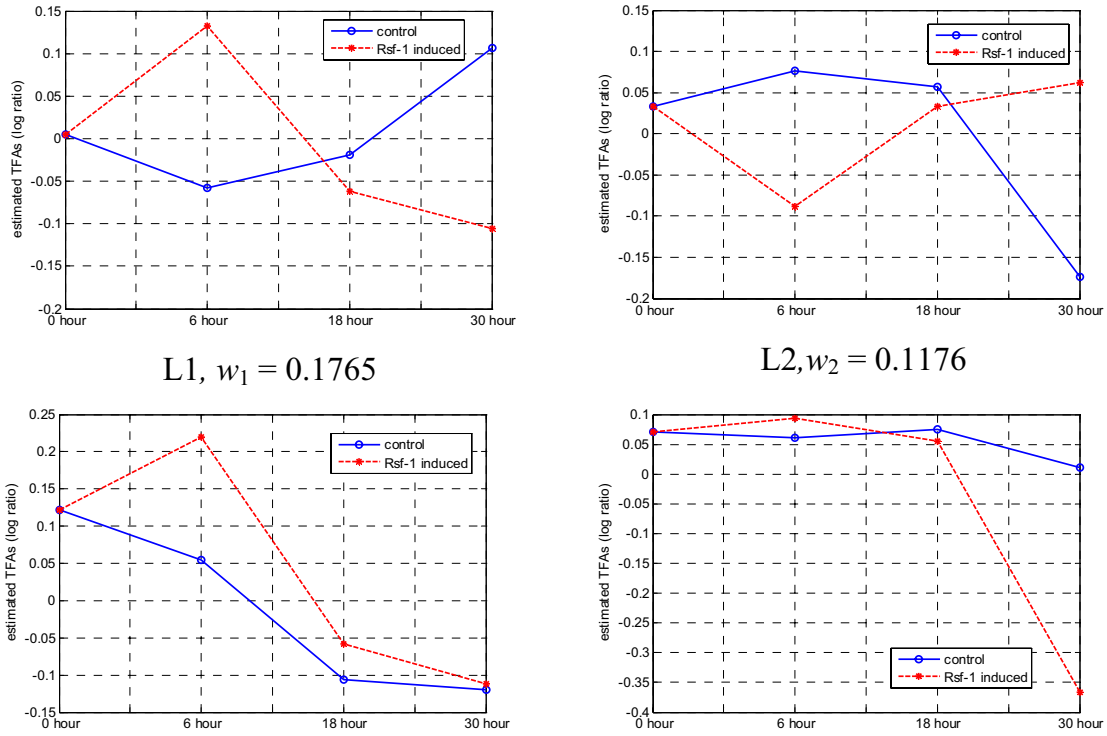


Figure 12
Estimated four knowledge-related TFAs using the proposed muti-scale ICA method. X-axis represents the time and Y-axis represents the estimated TFAs.

Table 5: Ovarian cancer-related TFs and their TRANSFAC entry IDs & descriptions

Index	TF Name	PWM Access No.	Consensus Binding Site	Factor Description
1	AP-2	M00189	MKCCCSCNGGCG	Activator protein 2
2	AP-2alpha	M00469	GCCNNNRGS	Activating enhancer binding protein 2 alpha
3	AP-2alphaA	M01045	ANNGCCTNAGGSNNT	Activating protein 2, AP-2A, Ker-1
4	AP-2gamma	M00470	GCCYNNGGS	Activator protein 2gamma, ERF-1
5	AP-2rep	M00933	CCCCGCCCCN	Specificity protein1, stimulating protein 1
6	BRCA1	M01082	KTNNGTTG	Breast cancer type 1 susceptibility protein
7	E2F	M00516	TTTSGCGCGMNR	EIIF protein, activator of myc, important for p107 promoter activity
8	Elk-1	M00007	NAAACMGGAAGTNCVH	Elk1, member of ETS oncogene family
9	NF-kappaB	M00774	NNNNKGGRAANTCCCN	Nuclear factor kappa B, p50
10	Sp1	M00933	CCCCGCCCCN	Specificity protein1, stimulating protein 1
11	TGIF	M00418	AGCTGTCANNA	5'-TG-3' interacting factor, TG-interacting factor, TGFB-induced factor
12	c-Rel	M00053	SGGRNTTTCC	Nuclear factor kappa B c-Rel, p68
13	P53	M00272	NGRCWTGYCY	Tumor protein p53, TRP53
14	ER	M00191	NNARGNCANNNTGACCYNN	Estrogen receptor

Discussion with biological interpretation

To enable a more detailed analysis, the top 10 genes extracted by optimal multi-scale ICA method are listed in Table 6 and the putative TFs in their promoter regions are shown in Fig. 14. Since none of the genes are in the KGP, they were entered into an Ingenuity Pathways Analysis (IPA) where we found that all of these genes can be incorporated into a single hypothetical network (Fig. 15). The major functions of this network are involved in gene expression, cancer development, and cellular motility. Five genes, FOSB, FOS, EGR1, IL8 and CDK2, are in the cancer module with p-values ranging from 1.84E-7 to 6.5E-3. FOSB and FOS belong to the Fos family that hetero-dimerizes with Jun proteins to form the AP-1 tran-

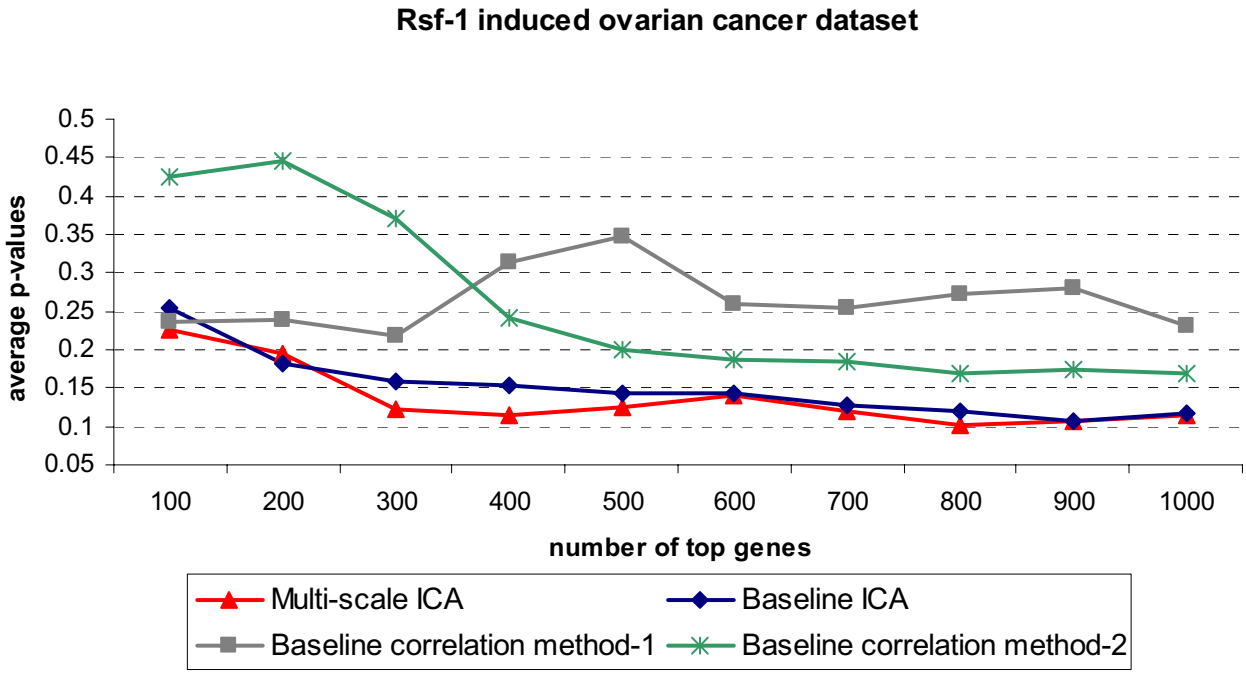


Figure 13
Average p-value of TF enrichment for different gene sets associated with different methods on Rsf-1-induced ovarian cancer microarray data set.

Table 6: Top 10 genes selected by the proposed multi-scale ICA on Rsf-1-induced microarray data

Rank	Probe Set ID	Gene Symbol	Gene Full Name
1	202768_at	FOSB	FBJ murine osteosarcoma viral oncogene homolog B
2	209189_at	FOS	v-fos FBJ murine osteosarcoma viral oncogene homolog
3	205476_at	CCL20	chemokine (C-C motif) ligand 20
4	212009_s_at	STIP1	stress-induced-phosphoprotein 1
5	209795_at	CD69	CD69 molecule
6	211506_s_at	IL8	interleukin 8
7	1557910_at	HSP90AB1	heat shock protein 90 kDa alpha (cytosolic), class B member 1
8	227404_s_at	EGR1	Early growth response 1
9	211804_s_at	CDK2	cyclin-dependent kinase 2
10	208621_s_at	VIL2	villin 2

scription factor complex [31]. AP-1 transcription factors control rapid responses of mammalian cells to stimuli that are associated with proliferation, differentiation and transformation [32]. IL-8 is a member of the C-X-C family of chemokines, and overexpression of IL-8 is observed in subsets of human ovarian cancer cells [33]. Previous studies have shown that the expression of interleukin-8 (IL-8) is directly correlated with the progression of human ovarian carcinomas implanted into the peritoneal cavity of nude mice [34]. The early growth response 1 (EGR1) is a transcription factor that acts as both tumor suppressor and tumor promoter depending on the cellular context. In the experiments of multiple pituitary and ovarian defects in Krox-24 (NGFI-A, Egr-1)-targeted mice, EGR1 was implicated as a novel key regulator of anterior pituitary physiology and that it may play important roles in specific cell lineages [35]. CDK2 is known to be involved in cell cycle regulation and the overexpression of CDK2 is associated with malignancy in ovarian tumors [36].

Conclusion

Biomarker identification is an important goal in many microarray data analyses. We propose a novel method, knowledge-guided multi-scale ICA, to find relevant biomarkers associated with specific biological functions. We aimed to infer knowledge-relevant regulatory signals and then identify corresponding biomarkers through a multi-scale strategy. A knowledge gene pool is constructed from multiple knowledge sources to help identify disease-specific gene clusters. By applying ICA to multi-scale gene clusters, an examination of the revealed regulatory modes can uncover knowledge of the underlying biological regu-

Gene/Promoter	-100	-200	-300	-400	-500	-600	-700	-800	-900	-1000	-1100	-1200	-1300	-1400	-1500	-1600	-1700	-1800	-1900	-2000
FOSB			Elk-1 Ap-2	E2F	Sp1 Ap-1									AP-2	ER E2F					
FOS		AP-2	E2F	AP-2 gamma	Elk-1			AP-2 gamma	E2F	Ap-2 gamma	Sp1			E2F	Ap-2 gamma	AP-2	E2F		BRCA1 Sp1	
CCL20			Elk-1			TGIF					ELK-1				TGIF		BRCA1			
STIP1	SP1	AP-2					BRCA1						Sp1	NF-kappaB						
CD69		c-Rel	c-Rel	NF-kappaB					BRCA1				AP-2 rep	E2F-1					BRCA1	
IL8															BRCA1			Elk-1	Elk-1	
HSP90AB1				AP-2						AP-2rep									Elk-1	AP-2rep NFKB
EGR1			E2F			Elk-1		AP-2		Elk-1		E2F-1		E2F		Sp1	Elk-1			AP-2 rep
CDK2	E2F								E2F	Sp1				AP-2alphaA						ER AP-2
EZR				AP-2rep	E2F-1	E2F		Sp1			E2F	AP-2					AP-2alphaA	AP-2rep		

Figure 14

TFs and their locations in 2 Kbp promoter region for top 10 genes selected by our approach. The promoter region is represented from -2,000 bp to 0 from TSS and each block in the figure represents a 100 bp region.

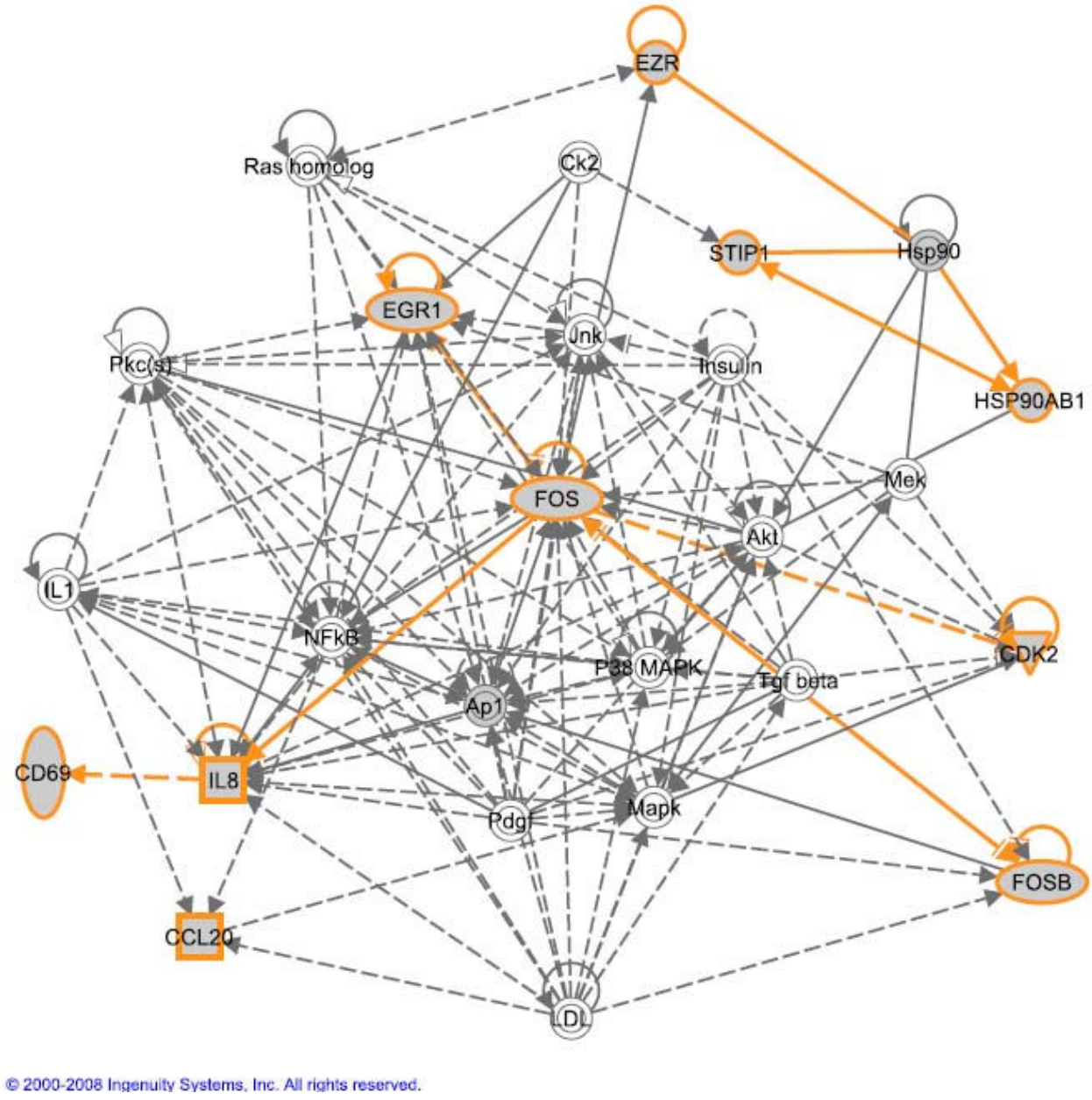


Figure 15
The network obtained from IPA with all of top 10 genes in Table 6. Five genes, FOSB, FOS, EGR1, IL8 and CDK2, are highly related to cancer module.

latory mechanisms. In addition, we have designed a statistical test procedure to measure the transcription factor enrichment of a selected gene set based on motif information. The approach was successfully applied to two gene expression profile data sets to identify biomarkers: yeast cell cycle microarray data and Rsf-1-induced microarray data. The experimental results show that our method can extract apparently biologically meaningful and condition-related biomarkers. The performance of the proposed method significantly outperforms several baseline methods for biomarker identification. More importantly, the proposed method has notable potential to discover novel biomarkers beyond any partial prior knowledge.

Authors' contributions

LC and JX formulated the problem and developed the theoretical framework of the algorithm. LC and CW carried out the development and implementation of the algorithm. IS and ZZ directed the application of the algorithm to the ovarian cancer data set. EH, RC and YW provided technical and biological support to the project. All authors participated in the writing of the manuscript, and have read and approved the manuscript.

Acknowledgements

This research was supported in part by NIH Grants (NS29525-13A, EB000830, CA109872, CA096483 and CA129080) and DoD/CDMRP Grant (BC030280). We also thank the anonymous reviews for their invaluable inputs that lead to several important improvements of this manuscript.

References

1. Devore J, Peck R: **Statistics: The Exploration and Analysis of Data.** CA Duxbury Press; 1997.
2. Tusher VG, Tibshirani R, Chu G: **Significance analysis of microarrays applied to the ionizing radiation response.** *Proc Natl Acad Sci USA* 2001, **98(9):**5116-5121.
3. Storey JD, Xiao W, Leek JT, Tompkins RG, Davis RW: **Significance analysis of time course microarray experiments.** *Proc Natl Acad Sci USA* 2005, **102(36):**12837-12842.
4. Conesa A, Nueda MJ, Ferrer A, Talon M: **maSigPro: a method to identify significantly differential expression profiles in time-course microarray experiments.** *Bioinformatics* 2006, **22(9):**1096-1102.
5. Hartigan JA, Wong MA: **A K-means clustering algorithm.** *App Statist* 1978, **28:**100-108.
6. Kohonen T: **Self-Organizing Maps.** NY: Springer; 1997.
7. Clarke R, Ressom HW, Wang A, Xuan J, Liu MC, Gehan EA, Wang Y: **The properties of high-dimensional data spaces: implications for exploring gene and protein expression data.** *Nat Rev Cancer* 2008, **8(1):**37-49.
8. Basso K, Margolin AA, Stolovitzky G, Klein U, Dalla-Favera R, Califano A: **Reverse engineering of regulatory networks in human B cells.** *Nat Genet* 2005, **37(4):**382-390.
9. Segal E, Shapira M, Regev A, Pe'er D, Botstein D, Koller D, Friedman N: **Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data.** *Nat Genet* 2003, **34(2):**166-176.
10. Liebermeister W: **Linear modes of gene expression determined by independent component analysis.** *Bioinformatics* 2002, **18(1):**51-60.
11. Hori G, Inoue M, Nishimura S, Nakahara H: **Blind gene classification on ICA of microarray data.** ICA: 2001; San Diego, CA; 2001:332-336.
12. Lee SI, Batzoglou S: **Application of independent component analysis to microarrays.** *Genome Biol* 2003, **4(11):**R76.
13. Saidi SA, Holland CM, Kreil DP, MacKay DJ, Charnock-Jones DS, Print CG, Smith SK: **Independent component analysis of microarray data in the study of endometrial cancer.** *Oncogene* 2004, **23(39):**6677-6683.
14. Hyvarinen A, Karhunen J, Oja E: **Independent Component Analysis.** John Wiley & Sons; 2001.
15. Gong T, Xuan J, Wang C, Li H, Hoffman E, Clarke R, Wang Y: **Gene module identification from microarray data using nonnegative independent component analysis.** *Gene Regulation and Systems Biology* 2007, **1:**349-363.
16. Wang C, Xuan J, Gong T, Clarke R, Hoffman E, Wang Y: **Stability based dimension estimation of ICA with application to microarray data analysis.** *The International Conference on Bioinformatics & Computational Biology: 2007* 2007.
17. Liao JC, Boscolo R, Yang YL, Tran LM, Sabatti C, Roychowdhury VP: **Network component analysis: reconstruction of regulatory signals in biological systems.** *Proc Natl Acad Sci USA* 2003, **100(26):**15522-15527.
18. Conlon EM, Liu XS, Lieb JD, Liu JS: **Integrating regulatory motif discovery and genome-wide expression analysis.** *Proc Natl Acad Sci USA* 2003, **100(6):**3339-3344.
19. Joung JG, Shin D, Seong RH, Zhang BT: **Identification of regulatory modules by co-clustering latent variable models: stem cell differentiation.** *Bioinformatics* 2006, **22(16):**2005-2011.
20. Wang C, Chen L, Zhao P, Hoffman E, Wang Y, Clarke R, Xuan J: **Motifdirected network component analysis for regulatory network inference.** *Sixth International Conference on Bioinformatics: 2007; Hong Kong, China* 2007.
21. Hyvarinen A, E O: **A fast fixed-point algorithm for independent component analysis.** *Neural Compuatation* 1997, **9:**1483-1492.
22. Frigyesi A, Veerla S, Lindgren D, Hoglund M: **Independent component analysis reveals new and biologically significant structures in micro array data.** *BMC Bioinformatics* 2006, **7:**290.
23. Matys V, Kel-Margoulis OV, Fricke E, Liebich I, Land S, Barre-Dirrie A, Reuter I, Chekmenev D, Krull M, Hornischer K, *et al.*: **TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes.** *Nucleic Acids Res* 2006:D108-110.
24. Karolchik D, Baertsch R, Diekhans M, Furey TS, Hinrichs A, Lu YT, Roskin KM, Schwartz M, Sugnet CW, Thomas DJ, *et al.*: **The UCSC Genome Browser Database.** *Nucleic Acids Res* 2003, **31(1):**51-54.
25. Kel AE, Gossling E, Reuter I, Cheremushkin E, Kel-Margoulis OV, Wingender E: **MATCH: A tool for searching transcription factor binding sites in DNA sequences.** *Nucleic Acids Res* 2003, **31(13):**3576-3579.
26. Witten I, Frank E: **Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations.** Morgan Kaufmann; 2000.
27. Spellman PT, Sherlock G, Zhang MQ, Iyer VR, Anders K, Eisen MB, Brown PO, Botstein D, Futcher B: **Comprehensive identification of cell cycle-regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization.** *Mol Biol Cell* 1998, **9(12):**3273-3297.
28. Shih Ie M, Sheu JJ, Santillan A, Nakayama K, Yen MJ, Bristow RE, Vang R, Parmigiani G, Kurman RJ, Trope CG, *et al.*: **Amplification of a chromatin remodeling gene, Rsf-1/HBXAP, in ovarian carcinoma.** *Proc Natl Acad Sci USA* 2005, **102(39):**14004-14009.
29. Affymetrix: **Guide to Probe Logarithmic Intensity Error (PLIER) Estimation.** Edited by: . Affymetrix I Santa Clara, CA; 2005.
30. Huang JY, Shen BJ, Tsai WH, Lee SC: **Functional interaction between nuclear matrix-associated HBXAP and NF-kappaB.** *Exp Cell Res* 2004, **298(1):**133-143.
31. Karin M-L: **The Fos family of transcription factors and their role in tumourigenesis, European journal of cancer.** *European journal of cancer* 2005, **41:**2449-2461.
32. Sharma SC, Richards JS: **Regulation of AP1 (Jun/Fos) factor expression and activation in ovarian granulosa cells. Relation of JunD and Fra2 to terminal differentiation.** *J Biol Chem* 2000, **275(43):**33718-33728.
33. Lee LF, Hellendall RP, Wang Y, Haskill JS, Mukaida N, Matsushima K, Ting JP: **IL-8 reduced tumorigenicity of human ovarian cancer in vivo due to neutrophil infiltration.** *J Immunol* 2000, **164(5):**2769-2775.
34. Xu L: **Ovarian cancer angiogenesis, biology and therapy.** University of Texas; 2000.
35. Topilko P, Schneider-Maunoury S, Levi G, Trembleau A, Gourdji D, Driancourt MA, Rao CV, Charnay P: **Multiple pituitary and ovarian defects in Krox-24 (NGFI-A, Egr-1)-targeted mice.** *Mol Endocrinol* 1998, **12(1):**107-122.
36. Hayami R, Sato K, Wu W, Nishikawa T, Hiroi J, Ohtani-Kaneko R, Fukuda M, Ohta T: **Down-regulation of BRCA1-BARD1 ubiquitin ligase by CDK2.** *Cancer Res* 2005, **65(1):**6-10.

ERRγ Mediates Tamoxifen Resistance in Novel Models of Invasive Lobular Breast Cancer

Rebecca B. Riggins,[1] Jennifer P-J. Lan,[1] Uwe Klimach,[1] Alan Zwart,[1] Luciane R. Cavalli,[1] Bassem R. Haddad,[1] Li Chen,[2] Ting Gong,[2] Jianhua Xuan,[2] Stephen P. Ethier,[3] and Robert Clarke[1]

[1]Lombardi Comprehensive Cancer Center and Department of Oncology, Georgetown University School of Medicine, Washington, District of Columbia; [2]Department of Electrical and Computer Engineering, Virginia Polytechnic Institute and State University, Arlington, Virginia; and [3]Barbara Ann Karmanos Cancer Institute and Department of Pathology, Wayne State University, Detroit, Michigan

Abstract

One-third of all estrogen receptor (ER)–positive breast tumors treated with endocrine therapy fail to respond, and the remainder is likely to relapse in the future. Almost all data on endocrine resistance has been obtained in models of invasive ductal carcinoma (IDC). However, invasive lobular carcinomas (ILC) comprise up to 15% of newly diagnosed invasive breast cancers each year and, whereas the incidence of IDC has remained relatively constant during the last 20 years, the prevalence of ILC continues to increase among postmenopausal women. We report a new model of Tamoxifen (TAM)-resistant invasive lobular breast carcinoma cells that provides novel insights into the molecular mechanisms of endocrine resistance. SUM44 cells express ER and are sensitive to the growth inhibitory effects of antiestrogens. Selection for resistance to 4-hydroxytamoxifen led to the development of the SUM44/LCCTam cell line, which exhibits decreased expression of ERα and increased expression of the estrogen-related receptor γ (ERRγ). Knockdown of ERRγ in SUM44/LCCTam cells by siRNA restores TAM sensitivity, and overexpression of ERRγ blocks the growth-inhibitory effects of TAM in SUM44 and MDA-MB-134 VI lobular breast cancer cells. ERRγ-driven transcription is also increased in SUM44/LCCTam, and inhibition of activator protein 1 (AP1) can restore or enhance TAM sensitivity. These data support a role for ERRγ/AP1 signaling in the development of TAM resistance and suggest that expression of ERRγ may be a marker of poor TAM response. [Cancer Res 2008;68(21):8908–17]

Introduction

Breast cancer is the second-most common cause of cancer-related death in women (1). One of the challenges in treating breast cancer is addressing the biological heterogeneity evident in the existence of several histologic and molecular subtypes. Two of the major histologic breast cancer classifications are invasive ductal carcinoma (IDC) and invasive lobular carcinoma (ILC). Currently, ILCs comprise up to 15% of invasive breast cancer diagnoses annually (2). Although the incidence of IDC has remained relatively constant during the last 20 years, a significant increase in ILC diagnosis is evident among postmenopausal women in Western Europe and the United States (reviewed in ref. 3). Although the increased use of estrogen plus progestin hormone replacement therapy for relief of perimenopausal and postmenopausal symptoms during this same time period may have contributed to the increase in ILC incidence (3), the precise mechanism(s) remains uncertain.

The clinical and pathologic features of lobular tumors are unique. ILC typically invades in a linear pattern, creating a longer, thinner mass, which is more difficult to detect by mammography, ultrasound, or breast self-exam (3). ILCs have a greater tendency to be bilateral, and women with this type of breast cancer are frequently older and have larger tumors at the time of their diagnosis (3). A higher incidence of ILC has been reported among women who initially present to the clinic with metastatic breast cancer (4). Although recent clinical studies imply that ILC is less responsive to neoadjuvant cytotoxic chemotherapy as a precursor to breast-conserving surgery (5, 6), there are conflicting reports as to whether patients diagnosed with ILC have a poorer, equivalent, or improved prognosis and overall survival when compared with IDC (reviewed in ref. 3).

Breast cancer patients whose tumors express estrogen receptor (ER) α (ERα) may be offered endocrine or antiestrogen therapy in addition to or in place of conventional chemotherapies. Currently, the most widely used antiestrogen is the triphenylethylene Tamoxifen (TAM), which functions as a partial antagonist by competing with estrogen for binding to the ER. TAM is known to induce a statistically significant improvement in the overall survival rate from breast cancer (7), and ~70% of all ER-positive (ER+)/progesterone receptor (PR)-positive (PR+) breast cancers will respond to TAM. When compared with IDC, a significantly greater percentage of ILC tumors are ER+/PR+ (discussed in ref. 3), suggesting that women diagnosed with this tumor subtype should be ideal candidates for endocrine therapy. However, study results differ as to whether ILC patients experience a better or worse risk of mortality than IDC patients after antiestrogen treatment (8, 9).

Regardless of tumor subtype, the development of endocrine resistance is a pervasive clinical problem (10–12). One-third of ER+/PR+ breast tumors treated with TAM do not respond to initial treatment, and the remaining 70% are still at risk to relapse in the future. A number of mechanisms have been proposed to control antiestrogen resistance in ER+ breast cancer (13), but many details of these mechanisms continue to be unclear. Studying endocrine resistance specifically in ILC has not been possible because of the lack of appropriate models; the most common models of resistance (notably MCF-7 cells) are derived from ductal adenocarcinomas (14).

Note: Supplementary data for this article are available at Cancer Research Online (http://cancerres.aacrjournals.org/).

Requests for reprints: Rebecca B. Riggins, Room E407 Research Building, Lombardi Comprehensive Cancer Center, Georgetown University School of Medicine, 3970 Reservoir Road Northwest, Washington, DC 20057. Phone: 202-687-7451; Fax: 202-687-7505; E-mail: rbr7@georgetown.edu.

doi:10.1158/0008-5472.CAN-08-2669

Given the unique clinical and molecular features of lobular tumors, and the suggestion that ILC tumors may respond less well to endocrine therapy, we have developed an ILC-specific cell culture model of endocrine resistance. The SUM44 breast cancer cell line was isolated from an ILC metastasis (15), is ER+/PR+, and displays other common features of ILC such as the loss of E-cadherin (16). We show that SUM44 cells contain functional ER and are sensitive to growth inhibition by antiestrogens. Selection of SUM44 cells against 4-hydroxytamoxifen (4HT) led to the establishment of the SUM44/LCCTam cell line, which is stably resistant to TAM. We then identified candidate genes associated with the endocrine resistant phenotype in SUM44/LCCTam cells and found changes in the expression of ERα and the estrogen-related receptor γ (ERRγ). Our mechanistic studies show that knockdown of ERRγ in the resistant cell line, and overexpression of ERRγ in endocrine-responsive lobular breast cancer cells, modulates TAM sensitivity. Finally, we show that ERRγ-driven transcription is increased in the resistant SUM44/LCCTam cell line, and inhibition of activator protein 1 (AP1) can restore or enhance TAM sensitivity in this model system.

Materials and Methods

Cell culture and reagents. All cells were shown to be free of *Mycoplasma* spp. contamination and maintained in a humidified incubator at 37°C in an atmosphere containing 95% air/5% CO_2. Routine tissue culture reagents (culture medium and additives, PBS, trypsin, etc.) were purchased from Invitrogen.

SUM44 cells were routinely cultured in serum-free medium plus insulin and hydrocortisone (SFIH) as described previously (15). LCCTam cells were maintained in SFIH containing 500 nmol/L 4HT (Sigma). LCCTam cells were cultured in SFIH in the absence of 4HT for 1 wk before all experiments. When SUM44 and LCCTam were passaged, cells were seeded in SFIH containing 2% fetal bovine serum (FBS) for the first 24 h to neutralize trypsin and promote cell attachment. MCF-7 cells were originally obtained from Dr. Marvin Rich (Karmanos Cancer Center, Detroit, MI), and MDA-MB-134 VI breast cancer cells were purchased from American Type Culture Collection; both were maintained in improved minimal essential medium with phenol red supplemented with 5% FBS.

17β-Estradiol (estradiol, E2) was purchased from Sigma; Fulvestrant (ICI 182,780; Fulv) and the c-JUN peptide inhibitor were purchased from Tocris Bioscience. The 3xERE-tk-luc promoter-reporter plasmid was kindly provided by Dr. Malcolm G. Parker (Imperial College, London, United Kingdom; ref. 17), 3xSF1RE-luciferase was a gift from Dr. Jean-Marc Vanacker (Institut de Génomique Fonctionnelle de Lyon, Université de Lyon, Lyon, France; ref. 18), and 3xAP1-luciferase was generously provided by Dr. Richard Pestell (Kimmel Cancer Center, Thomas Jefferson University, Philadelphia, PA). The plasmid encoding wild-type murine ERRγ bearing an NH_2-terminal hemagglutinin (HA) tag (pSG5-HA-ERR3) was a gift from Dr. Michael Stallcup (Keck School of Medicine, University of Southern California, Los Angeles, CA; ref. 19). Small inhibitory RNA (siRNA) oligonucleotide duplexes directed against ERRγ (siGENOME SMARTpool), nonsilencing control oligonucleotides, and the DharmaFECT 1 reagent were purchased from Dharmacon. The FuGene 6 transfection reagent was purchased from Roche.

Luciferase promoter-reporter assays. Cells were seeded in SFIH at a density of 9×10^4 cells per well in 12-well plastic tissue culture dishes for 24 to 48 h before transfection with 0.6 μg luciferase promoter-reporter construct and 0.2 μg phRL-SV40 Renilla internal control (Promega). The following day, transfected cells were refed with SFIH, or SFIH containing 10 nmol/L E2, 1,000 nmol/L 4HT, 100 nmol/L Fulv, 20 μmol/L c-JUN peptide inhibitor, or ethanol vehicle as indicated in each figure for a further 24 h before lysis and measurement of luciferase activity by using the Dual Luciferase Assay kit (Promega) as described previously (20). Luminescence was quantified using a Lumat LB 9501 luminometer (EG&G Berthold).

Proliferation assays. Cells were seeded in SFIH at a density of 2 to 3×10^4 per well in 24-well plastic tissue culture dishes 1 d before the addition of the indicated concentrations of drug or ethanol vehicle. Cells were cultured for 6 d with two medium changes before being trypsinized, resuspended in PBS, and counted using a Z1 Single Coulter Counter (Beckman/Coulter). At least three independent assays were performed in triplicate or quadruplicate, and the data were normalized to vehicle-treated cells.

BrdUrd ELISAs. Cells were seeded in SFIH at a density of 1×10^4 cells per well in 96-well plastic tissue culture dishes 1 d before the addition of drug or ethanol vehicle as indicated. Cells were then cultured for ~54 h before the addition of BrdUrd (final concentration 10 μmol/L) for an additional 18 h (total incubation in drug, 72 h) before performing the Cell Proliferation ELISA, BrdUrd (colorimetric) assay as directed by the manufacturer (Roche). At least three independent assays were performed with five replicate wells per treatment group, and data were normalized to vehicle-treated cells.

BrdUrd immunofluorescence assays. These assays were performed as described above (drug treatment and BrdUrd addition) and by Riggins and colleagues (cell seeding and staining procedures; ref. 21) with the following modifications: ERRγ expression was detected using the HA.11 monoclonal antibody from Covance (1:500) followed by AlexaFluor594-conjugated goat anti-mouse secondary antibody (Invitrogen; 1:500), and BrdUrd incorporation was detected using the AlexaFluor488-conjugated anti-BrdUrd antibody (1:10; BD Biosciences). Cells were visualized on a Nikon E600 epifluorescence microscope at ×20 magnification.

Cell cycle analysis. Cells were seeded in SFIH at a density of 5×10^4 cells per well in 6-well plastic tissue culture dishes 1 d before the addition of 1,000 nmol/L 4HT or ethanol vehicle. Cells were then cultured for 72 h before harvesting and cell cycle analysis by the Vindelov method (22).

Derivation of SUM44/LCCTam cells. A TAM-resistant SUM44 variant was established according to previously published procedures (23). Subconfluent T-25 cm^2 tissue culture flasks of SUM44 cells were selected against increasing concentrations of 4HT, beginning with 1 nmol/L. After 3 passages of the cells at each dose, the drug concentration was increased (1→5→10→50→100→500 nmol/L), terminating at a concentration of 500 nmol/L 4HT. Cells proliferating in 500 nmol/L 4HT were designated SUM44/LCCTam (hereafter abbreviated as LCCTam). LCCTam cells were cultured in SFIH in the absence of 4HT for 1 wk before all experiments.

Comparative genomic hybridization. Normal control DNA was prepared from peripheral blood lymphocytes of a normal donor and test DNA was extracted from the cultured cell lines (SUM44 and the TAM-resistant LCCTam variant) using standard protocols, and comparative genomic hybridization (CGH) was performed as previously described (24). Gray scale images from at least 10 metaphases from each hybridization were acquired with a cooled charge-coupled device CCD camera (CH250; Photometrics) connected to a Leica DMRBE microscope equipped with fluorochrome specific optical filters TR1, TR2, TR3 (Chroma Technology). Quantitative evaluation of the hybridization was done using commercially available software (Applied Imaging). Average ratio profiles were calculated as the mean value of at least eight ratio images to identify chromosomal copy number changes in all cases (see Supplementary Fig. S1).

RNA isolation, gene expression microarray preprocessing, and data analysis. Total RNA was extracted from subconfluent T-25 cm^2 tissue culture flasks of SUM44 and LCCTam cells, then processed and arrayed as described by Gomez and colleagues (25). Microarray data quality was then assessed using several tools, including those recommended by Affymetrix and a series of additional QC measures under development in our laboratory (26). The Robust Multiple-Array Average method was used to preprocess the raw gene expression data, as implemented in the Bioconductor project.[4] We then isolated a reduced dimension data set that included genes that exhibit ≥2 fold change ($P < 0.05$) and genes with intensity ≥log2 (10) in both SUM44 and SUM44/LCCTam groups. Data visualization before and after dimensionality reduction was facilitated by multidimensional scaling as estimated using Principal Component Analysis

[4] http://bioconductor.org

(PCA) and Discriminant Component Analysis (27), to ensure that the global structure of the data were not altered by dimensionality reduction procedures (see Supplementary Fig. S2). Expression data are available through the Gene Expression Omnibus database, accession GSE12708.

Real-time qPCR. Total RNA from independent cultures (not RNA from cultures used for microarray analysis) was isolated, cleaned, quantified, and reverse-transcribed as described in (25). qPCR reactions for each cDNA sample and a standard curve were performed using TaqMan Universal PCR Master Mix and the following TaqMan Gene Expression Assay primers (Applied Biosystems): ESR1, Hs00174860_m1; ESRRG, Hs00155006_m1; and the housekeeping gene RPLP0 (Hs99999902_m1) as in Gomez and colleagues (25). Expression data for each gene was estimated relative to the housekeeping control, and these data were used to calculate the ratio of expression relative to that in the parental SUM44 cell line.

Cell lysis and Western blot analysis. Subconfluent monolayers of cells were harvested, lysed, and analyzed by Western blot as in Bouker and colleagues (28). Primary antibodies for ERRγ (1:1,000), ERRα (1:500), and ERRβ (1:500) were purchased from GenWay. Antibodies for ERα (1:500) and ERβ (1:1,000) were purchased from NovoCastra and Affinity Bioreagents, respectively. Antibodies for FASN (1:500) and HMGCS2 (1:2,000) were purchased from Abcam. To confirm equal loading, membranes were reprobed using a β-actin monoclonal antibody (1:5,000) purchased from Sigma, or a glyceraldehyde-3-phosphate dehydrogenase (GAPDH) goat polyclonal antibody (1:5,000) purchased from Santa Cruz Biotechnology. Secondary antibodies conjugated to horseradish peroxidase were purchased from GE Healthcare and Santa Cruz Biotechnology. Densitometry was performed using NIH ImageJ software[5] and images were compiled using Adobe Photoshop CS2.

ERRγ siRNA. LCCTam cells were seeded in 96-well plastic tissue culture dishes in SFIH at 1×10^4 cells per well 1 d before transfection with 100 nmol/L ERRγ (siERRγ) or nonsilencing control siRNA oligonucleotides (siC) using DharmaFECT1 (Dharmacon) according to manufacturer's specifications. The next day, cells were treated with 1,000 nmol/L 4HT or ethanol vehicle before addition of BrdUrd for an additional 18 h (total incubation in drug, 48 h). Cell Proliferation ELISA, BrdUrd (colorimetric) assays were performed as described above. In parallel, cells were seeded in 12-well dishes at a density of 9×10^4 cells per well, transfected with 100 nmol/L siC or siERRγ, and cells were lysed on the same day that BrdUrd ELISAs were performed (total transfection time, 72 h) for Western blot analysis.

Statistics. All statistical calculations were performed using SigmaStat version 3.0 (Systat). Luciferase promoter-reporter, cell proliferation, BrdUrd, real-time reverse transcription-PCR (RT-PCR), and microarray data from *in vitro* studies were compared using either Student's *t* test or one-way ANOVA followed by *post hoc t* test, as appropriate, and indicated in the text and figure legends. Statistical significance is defined at ≥95% confidence level, or a *P* value of ≤0.05.

Results

SUM44 cells have functional ER and are sensitive to growth inhibition by 4HT. The SUM44 breast cancer cell line was derived from an ILC and a high percentage of ILC tumors are ER+ (29). Although this cell line is also ER+ (15), ER functional status is unknown and SUM44 responsiveness to estrogens and antiestrogens has not previously been determined. Therefore, SUM44 cells were transfected with the 3xERE-tk-luc reporter construct and stimulated with estrogen, antiestrogen, or ethanol control (Fig. 1*A*). Estrogen (E2) modestly but significantly induces, whereas 4HT significantly decreases, ERE-luciferase activity ($P < 0.001$). We also observed that the steroidal antiestrogen Fulv decreases ERE-luciferase activity, and that both 4HT and Fulv block the E2-induced stimulation of ERE-luciferase activity ($P < 0.001$). These data suggest that the SUM44 ER responds appropriately to estrogenic and antiestrogenic stimuli.

To determine whether SUM44 cells are sensitive to growth inhibition by 4HT, cells were treated with antiestrogen as indicated for 6 d (Fig. 1*B, closed circles*). 4HT significantly inhibits the proliferation of SUM44 cells (ANOVA $P < 0.001$). The observed reduction in cell number is also reflected in an inhibition of DNA synthesis as shown by reduced BrdUrd incorporation after 72 hours of 4HT treatment (ANOVA $P < 0.001$; Fig. 1*C, closed circles*), consistent with the known cytostatic effect of 4HT (12).

Generation of a TAM-resistant SUM44 variant. Because ILCs are predominantly ER+ and TAM has been the most widely used endocrine agent for the treatment of ER+ breast cancer, we sought to develop a TAM-resistant ILC model using the SUM44 cell line. Cells were selected against increasing concentrations of 4HT, and the cell population proliferating in 500 nmol/L 4HT (within the range of clinically relevant concentrations; ref. 10) was designated SUM44/LCCTam (hereafter called LCCTam).

The basal growth rate of LCCTam is identical to that of the parental SUM44 cell line and as expected, LCCTam cells are no longer responsive to the antiproliferative effects of 4HT (Fig. 1*B, open triangles, N.S.*), and LCCTam DNA synthesis is no longer inhibited by 4HT (ANOVA $P = 0.212$; Fig. 1*C, open triangles*). To further confirm that differences in SUM44 and LCCTam cell proliferation in response to antiestrogen reflect changes in sensitivity to the cytostatic effects of 4HT, we performed cell cycle analysis. SUM44 cells treated with 1 μmol/L 4HT show a significantly greater fraction of cells arrested in the G_1 phase compared with ethanol-treated controls ($P \leq 0.001$; data not shown), whereas 4HT no longer induces an accumulation of LCCTam cells in G_1 ($P = 0.722$, data not shown). Together, these findings show that SUM44 cell growth and cell cycle progression are efficiently inhibited by 4HT, but that LCCTam cells have acquired resistance to the inhibitory effects of this antiestrogen.

Changes in the transcriptome of LCCTam cells are not associated with chromosomal aberrations. To characterize further this novel ILC cell model, we determined the pattern of, and differences in, genomic alterations and gene expression between SUM44 and LCCTam cells using CGH and Affymetrix gene expression microarray analysis, respectively. The genetic lineage of the two cell lines was confirmed to be identical by DNA fingerprinting using genetic markers at nine different loci. CGH analysis revealed changes in the DNA copy number (gains, losses, and amplifications) in both SUM44 and LCCTam (Supplementary Fig. S1). Importantly, a comparison between our CGH findings and a previously reported CGH analysis of SUM44 show a similar pattern of aberrations (30). We found no significant difference in the pattern of chromosomal alterations between the two cell lines; acquired estrogen independence also is not associated with changes in the amplification of DNA sequences (31).

In marked contrast, microarray analysis reveals a large number of changes in gene expression. We used PCA (27) to visualize the high-dimensional data set in two dimensions; SUM44 and LCCTam are linearly separable in this two-dimensional PCA projection based on the top two principal components that capture 95% of the cumulative variance in the data (Supplementary Fig. S2). Using a final cutoff of ≥2-fold change with $P \leq 0.05$ (univariate, uncorrected, T-statistic), we find that 380 genes are likely to be significantly altered: expression of 91 genes are increased and 289 genes are decreased in LCCTam versus SUM44 controls (Supplementary Table S1).

To maintain focus on the TAM-resistant phenotype observed in LCCTam cells, we first chose to investigate gene expression

[5] http://rsb.info.nih.gov/ij

changes in ERs and other members of the nuclear receptor superfamily. Expression of ERα (HUGO symbol ESR1) is decreased 3.1-fold in LCCTam compared with SUM44 cells by microarray (P = 0.0013), which was subsequently confirmed by qPCR analysis (↓2.98-fold, P < 0.001; Fig. 2*A, white bars*). In contrast, expression of the orphan nuclear receptor ERRγ (HUGO symbol ESRRG) is 4.4-fold increased in the resistant LCCTam cells by microarray (P = 0.01) and 10-fold increased by qPCR (P = 0.03; Fig. 2*A, black bars*).

To confirm that differences in the mRNA expression of these receptors are maintained at the protein level, cell lysates were collected and analyzed for ERRγ and ERα expression by Western blot (Fig. 2*B, inset*). As observed for mRNA, ERRγ protein expression is increased (↑2.5-fold; P = 0.03) and ERα expression is decreased (↓2-fold; P = 0.03) in LCCTam cells. We also examined the protein levels of all other ERs and ERRs (ERβ, ERRα, and ERRβ) and find no differences in their expression between SUM44 and LCCTam cells (Fig. 2*C*).

ERRγ plays a functional role in TAM resistance in LCCTam cells. ERRγ is an orphan nuclear receptor with no known natural ligand that has been shown to have constitutive transcriptional activity at several DNA response elements (reviewed in refs. 32, 33). ERRγ and its family members ERRα1 and ERRβ bear some structural similarity to the ER (32, 34). Although ERRα1 has previously been shown to activate or repress estrogen response element (ERE)-mediated transcription depending on cellular context (34) and to participate in HER2-dependent signaling in BT474 breast cancer cells (35), the role of ERRγ in breast cancer therapeutic response is underexplored (36).

We hypothesized that if increased expression of ERRγ in LCCTam cells performs a functional role in the acquired TAM resistance phenotype, knockdown of receptor expression should restore TAM sensitivity. LCCTam cells were transiently transfected with siRNA oligonucleotides directed against ERRγ (siERRγ) or a nonsilencing control (siC) before treating the cells with 4HT and assessing DNA synthesis as measured by BrdUrd incorporation. A 2- to 3-fold decrease in ERRγ expression is attained by siRNA (P < 0.001; Fig. 3*A*). Importantly, ERRγ knockdown also partially restores sensitivity to 4HT in the LCCTam cells (P = 0.03 versus siERRγ ethanol and P < 0.001 versus siC in 1,000 nmol/L 4HT; Fig. 3*B*) with no effect on the expression of ERα (Fig. 3*A, inset, bottom*). These data suggest that ERRγ plays a key functional role in the LCCTam TAM resistance phenotype.

Overexpression of ERRγ induces 4HT resistance. Next, we sought to determine whether ERRγ overexpression could induce TAM resistance in endocrine-responsive breast cancer cells. SUM44 cells grown on fibronectin-coated coverslips were transiently transfected with the pSG5-HA-ERR3 plasmid, encoding the murine homologue of ERRγ, which is 100% identical to human ERRγ at the amino acid level (19), or the empty vector (pSG5). Cells were then treated with 1 μmol/L 4HT or ethanol vehicle and immunostained for BrdUrd incorporation (*green*) and ERRγ expression (HA, *red*; ref. 21). In agreement with our results in Fig. 1*C*, 4HT significantly

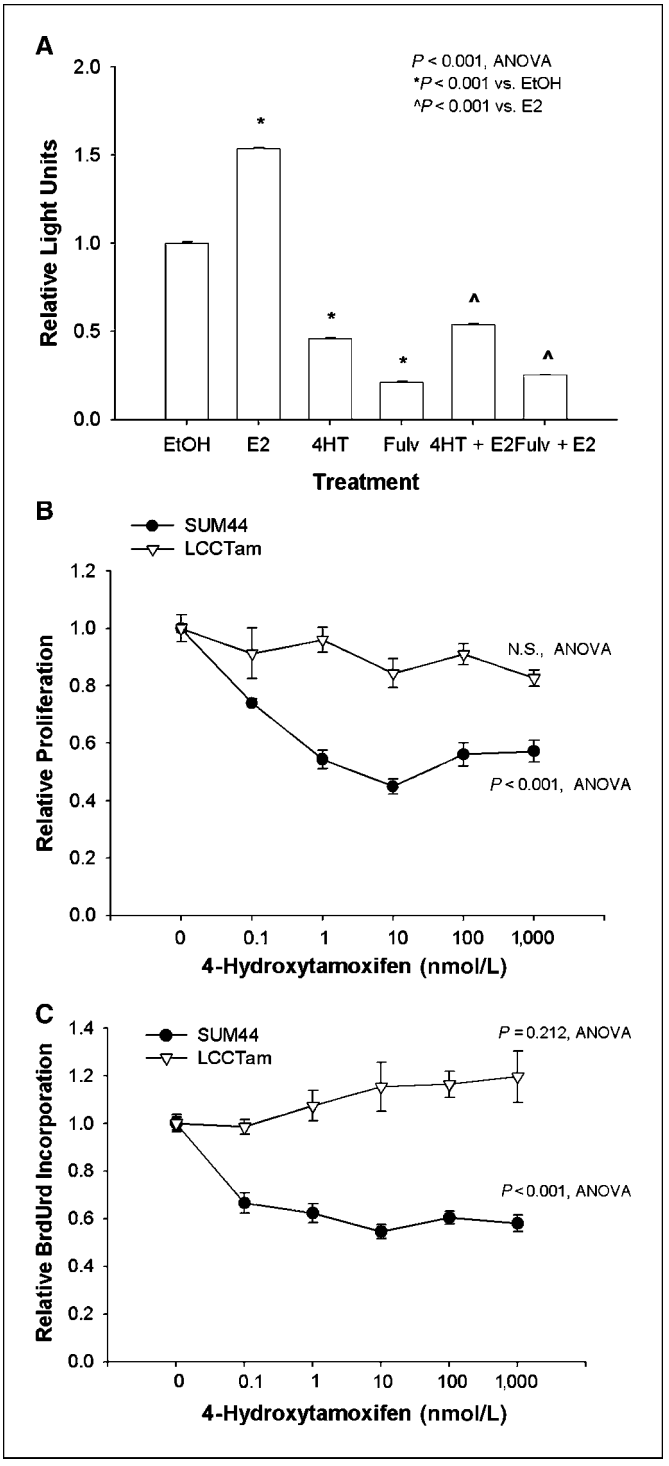


Figure 1. SUM44 cell proliferation and ER transcriptional activity are inhibited by antiestrogens, and LCCTam cells have acquired resistance to TAM. *A,* cells were seeded in 12-well tissue culture dishes, transfected with plasmids encoding 3xERE-tk-luciferase and phRL-SV40 Renilla, and treated with 10 nmol/L E2, 1,000 nmol/L 4HT, 100 nmol/L Fulv, 4HT+E2, Fulv+E2, or ethanol control for 24 h before harvest and luciferase assay. ERE-luciferase values are normalized to Renilla activity to obtain Relative Light Units, and data are presented as the mean relative to ethanol ± SE for a representative experiment performed in triplicate. ANOVA P < 0.001; *, P < 0.001 for comparisons to ethanol and ^, P < 0.001 for comparisons to E2 by *post hoc* Student's t test. *B,* cells were seeded in 24-well tissue culture dishes and treated with the indicated concentrations of 4HT for 6 d, at which time cell number was determined. *Points,* mean proliferation relative to ethanol for a representative experiment performed in quadruplicate; *bars,* SE. ANOVA P < 0.001 for SUM44, and not significant (*N.S.*) for LCCTam. *C,* cells were seeded in 96-well tissue culture dishes 1 d before treatment with the indicated concentrations of 4HT for a total of 72 h; BrdUrd was added for the last 18 h of culture. *Points,* mean BrdUrd incorporation relative to ethanol for a representative experiment performed in quintuplicate; *bars,* SE. ANOVA P < 0.001 for SUM44, and P = 0.212 (*N.S.*) for LCCTam.

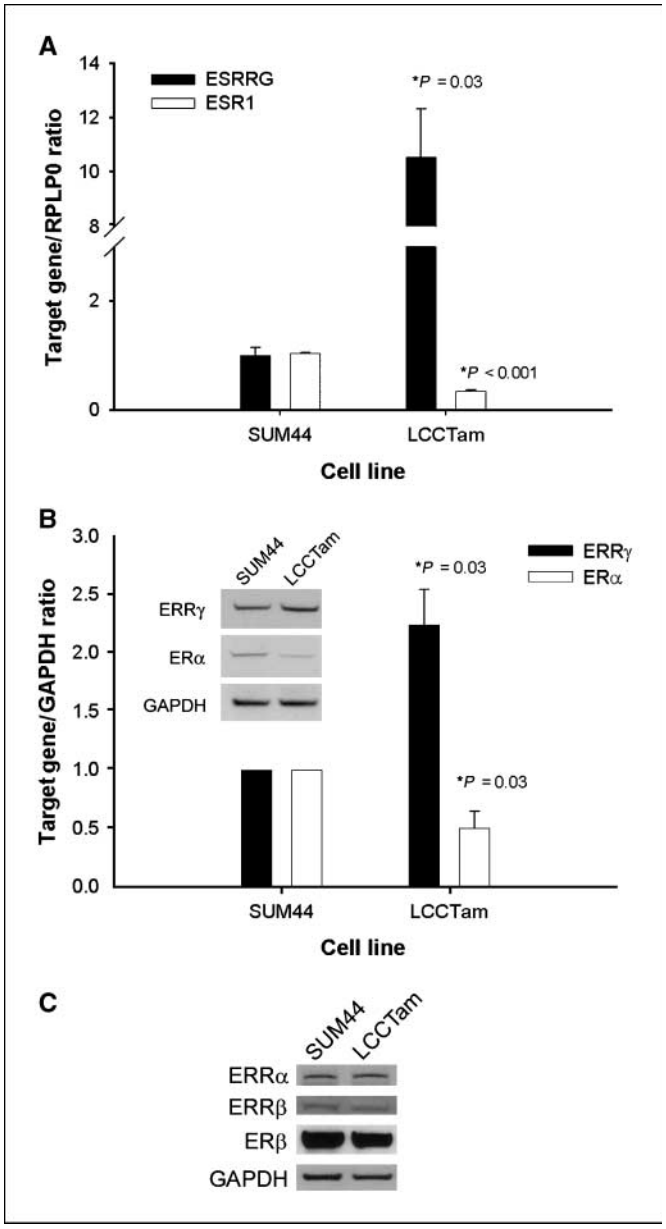


Figure 2. ERα and ERRγ mRNA and protein expression are significantly altered during the acquisition of TAM resistance. *A,* total RNA was isolated from SUM44 and LCCTam cells, reverse-transcribed, and subjected to RT-PCR to detect ERRγ (HUGO gene symbol ESRRG), ERα (*ESR1*), and the housekeeping gene RPLP0. *Columns,* mean target gene/RPLP0 ratio for three samples analyzed in triplicate; *bars,* SE. *P = 0.03 for ESRRG and *P < 0.001 for ESR1 in SUM44 versus LCCTam by Student's *t* test. *B,* densitometric quantification of protein expression and a representative Western blot are shown for expression of ERRγ, ERα, and the GAPDH loading control. *P = 0.03 for ERRγ and ERα in SUM44 versus LCCTam by Student's *t* test. *C,* representative Western blot showing ERRα, ERRβ, and ERβ expression, and the GAPDH loading control, in SUM44 and LCCTam cells.

reduces BrdUrd incorporation in SUM44 cells transfected with the empty vector pSG5 (P < 0.001; Fig. 4*A, ii* versus *iv*, 48.9% versus 16.9% BrdUrd incorporation). However, 4HT can no longer inhibit DNA synthesis when ERRγ is overexpressed (P < 0.001; Fig. 4*A, iv* versus *viii*, 16.9% versus 53.9% BrdUrd incorporation). The effect of ERRγ overexpression is particularly striking when comparing BrdUrd incorporation in transfected versus untransfected cells in the presence of 4HT within the same field of view. In Fig. 4*A*, most ERRγ-positive (*red*) cells incorporate BrdUrd (*viii, arrowheads*), whereas ERRγ-negative cells show little-to-no BrdUrd incorporation (*viii*, *).

To confirm that ERRγ can regulate TAM resistance in breast cancer cell lines other than SUM44, we performed the same study in MDA-MB-134 VI cells, which are ER+ and TAM-sensitive (37) and are also considered to be of lobular origin (38). When transfected with the pSG5 empty vector, DNA synthesis in MDA-MB-134 VI cells is inhibited by 4HT by nearly 2-fold (49.9% versus 27.3% BrdUrd incorporation, P < 0.001; Fig. 4*B*). However, when ERRγ is overexpressed, these cells become significantly less responsive to the inhibitory effects of 4HT (27.3% versus 44.7% BrdUrd incorporation, P = 0.001; Fig. 4*B*). Together, these data show that increased expression of ERRγ can induce TAM resistance in several ER+ lobular breast cancer cell lines.

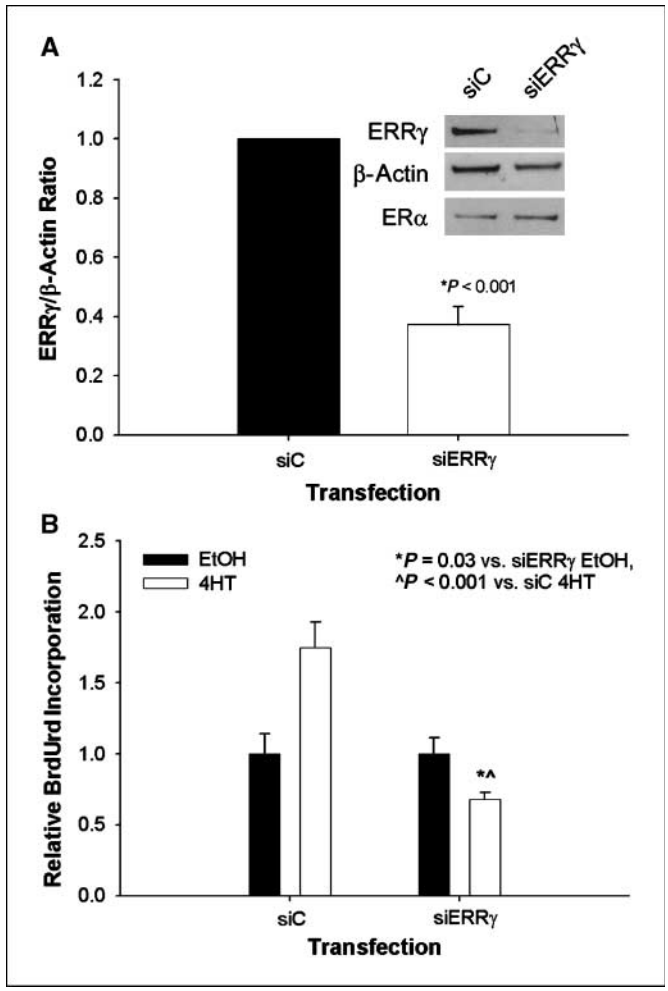


Figure 3. siRNA knockdown of ERRγ in LCCTam cells restores TAM sensitivity. *A,* cells seeded in 12-well dishes were transfected with control (*siC*) or ERRγ-specific (*siERRγ*) oligonucleotides (final concentration, 100 nmol/L) for 72 h before lysis, Western blot analysis, and densitometry. *Columns,* mean ERRγ/β-actin ratio for three independent experiments; *bars,* SE. *Inset,* a representative image. *P < 0.001 for siERRγ versus siC, Student's *t* test. ERRγ knockdown has no effect on ERα expression (*inset, bottom*). *B,* cells seeded in 96-well dishes were transfected with siC or siERRγ oligonucleotides 24 h before treatment with 1,000 nmol/L 4HT or ethanol control. BrdUrd assays were performed as described above; *columns,* mean BrdUrd incorporation relative to ethanol for a representative experiment performed in quintuplicate; *bars,* SE. *P = 0.03 versus siERRγ ethanol, and ^P = <0.001 versus siC in 1,000 nmol/L 4HT by Student's *t* test.

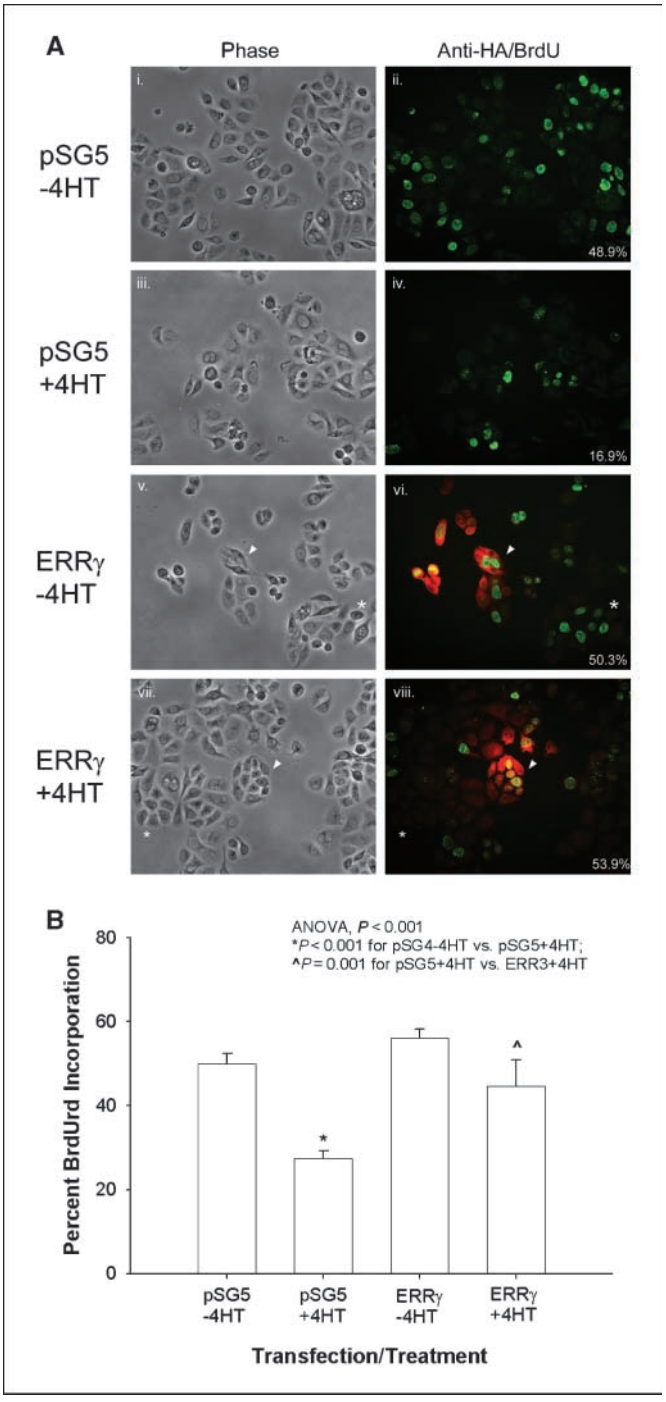


Figure 4. ERRγ overexpression in SUM44 and MDA-MB-134 VI breast cancer cells induces TAM resistance. *A,* SUM44 cells were seeded on fibronectin-coated coverslips, then transfected with pSG5-HA-ERR3 (*ERRγ*) or empty vector (*pSG5*). Cells were treated with 1,000 nmol/L 4HT or ethanol control for a total of 48 h; BrdUrd was added for the last 18 h of culture before fixation and processing for HA (*ERRγ*) and BrdUrd immunostaining. Representative phase-contrast and fluorescent images are shown (*red,* ERRγ; *green,* BrdUrd); *arrowheads,* ERRγ-positive cells; *, ERRγ-negative cells. Quantitative data are presented as the mean percent BrdUrd incorporation for a representative experiment in which 4 to 5 microscopic fields (>500 total cells) were counted per condition. ANOVA $P < 0.001$; $P < 0.001$ for pSG5-4HT versus pSG5+4HT, and $P < 0.001$ for pSG5+4HT versus ERRγ+4HT by *post hoc* Student's *t* test. *B,* MDA-MB-134 VI cells were seeded, transfected, drug treated, and stained as described in *A. Columns,* mean percent BrdUrd incorporation for a representative experiment in which 4 to 5 fields (and >500 total cells) per condition were counted; *bars,* SE. ANOVA $P < 0.001$; *, $P < 0.001$ for pSG5-4HT versus pSG5+4HT, and ^, $P = 0.001$ for pSG5+4HT versus ERRγ+4HT by *post hoc* Student's *t* test.

ERRγ-associated transcriptional activity is increased in resistant LCCTam cells. A crucial difference between ERRγ and liganded nuclear receptors like ERα is the regulation of their transcriptional activities. Whereas ERα is dependent on ligand for full activation, ERRγ and the other members of this orphan family exhibit constitutive transcriptional activity. The ERRγ DNA binding domain is ~64% identical to that of ERα (34). Consequently ERRγ can bind to the same EREs as ERα, but it can also potently activate the steroidogenic factor-1 response element (SF1RE; ref. 32). Although none of the ERR family members are affected by E2 stimulation because their ligand binding domains cannot accommodate E2 binding (discussed in ref. 34), ERRγ transcriptional activity at EREs and SF1REs can be inhibited by 4HT (39, 40). In contrast, 4HT-bound ERRγ acquires the ability to positively regulate transcription at AP1 sites (reviewed in ref. 34).

To begin to understand the mechanism by which ERRγ up-regulation confers resistance to LCCTam cells, we examined the activity of ERE-, SF1RE-, and AP1-driven luciferase promoter-reporter constructs transiently expressed in SUM44 and LCCTam cells (Fig. 5*A*). Luciferase expression controlled by the ERE and SF1RE response elements is significantly increased by 5- and 3-fold, respectively, in LCCTam cells compared with SUM44 cells ($P < 0.005$). When LCCTam cells are cultured in 4HT ("LCCTam+4HT"), ERE-luciferase activity is somewhat reduced but still shows a nearly 2-fold increase relative to SUM44 (*black bars*; $P < 0.005$), whereas SF1RE-luciferase activity remains high (*white bars*; 3-fold above the levels in SUM44 cells; $P < 0.005$). In contrast, AP1-luciferase activity increases up to 8-fold that observed in SUM44 cells in the presence of 4HT (*hatched bars*; $P < 0.005$).

ERRγ/AP1 activity seems to drive TAM resistance in LCCTam cells. To test whether the observed robust AP1 activity plays a functional role in the TAM-resistant phenotype, we used a cell-permeable peptide fragment of c-JUN that blocks its interaction with the JUN NH_2-terminal protein kinase, resulting in strong AP1 inhibition (41). This c-JUN peptide has virtually no effect on SF1RE-luciferase activity (Fig. 5*B, N.S.*) but can inhibit AP1-luciferase activity by >2-fold ($P = 0.04$; Fig. 5*C*). Importantly, this level of AP1 inhibition restores 4HT-mediated growth inhibition to LCCTam cells ($P = 0.001$; Fig. 5*D*) and enhances the sensitivity of the parental SUM44 cells to the growth-inhibitory effects of 4HT ($P = 0.002$).

Our functional data suggest that in LCCTam cells, increased ERRγ-driven AP1 transcriptional activity is most strongly associated with TAM resistance. However, endogenous ERRγ/AP1 target genes have yet to be identified; ERRγ-dependent AP1 activity has previously been reported only on heterologous promoter constructs (42). We therefore used the TRANSFAC Professional 11.1 database (43) to search the proximal promoter regions of genes up-regulated ≥2-fold in LCCTam cells for AP1 consensus sites within 5,000 bp of the start site. The MatchTM algorithm (44) was used to analyze the DNA sequences and search for potential AP1 binding sites, using Position Weight Matrices to minimize false positives. Several genes had multiple AP1 response elements in their promoter regions (Fig. 6*A*). Western blot analysis was then used to confirm the overexpression of two of these genes, HMGCS2 and FASN (Fig. 6*B*). HMGCS2 is a nuclear-encoded mitochondrial matrix gene that can regulate ketogenesis and cholesterol synthesis (45, 46), and FASN is the final enzyme of the fatty acid biosynthetic pathway (47). Components of all three processes have been

implicated in the etiology or progression of breast cancer, and FASN activity can affect hormonal sensitivity in breast and endometrial cancer cells (48–50). Therefore, we suggest that HMGCS2 and FASN may be two novel ERRγ/AP1 targets in TAM-resistant breast cancer.

Discussion

In this study, we report the development of the first model of endocrine-resistant breast cancer in a cell line derived from an invasive lobular breast carcinoma, and show that the orphan nuclear receptor ERRγ and its ability to drive AP1 transcriptional activity are central to the TAM resistance phenotype.

Selection of SUM44 cells against 4HT led to the establishment of the LCCTam cell line, which is stably resistant to TAM. In the resistant LCCTam cells, we observe a significant down-regulation of ERα (although they remain ER+), accompanied by a significant increase in the expression of ERRγ. Resistance to antiestrogens has been hypothesized to take place through several diverse mechanisms (10, 12). One is loss or mutation of ERα, whereas others include alterations in the profile of

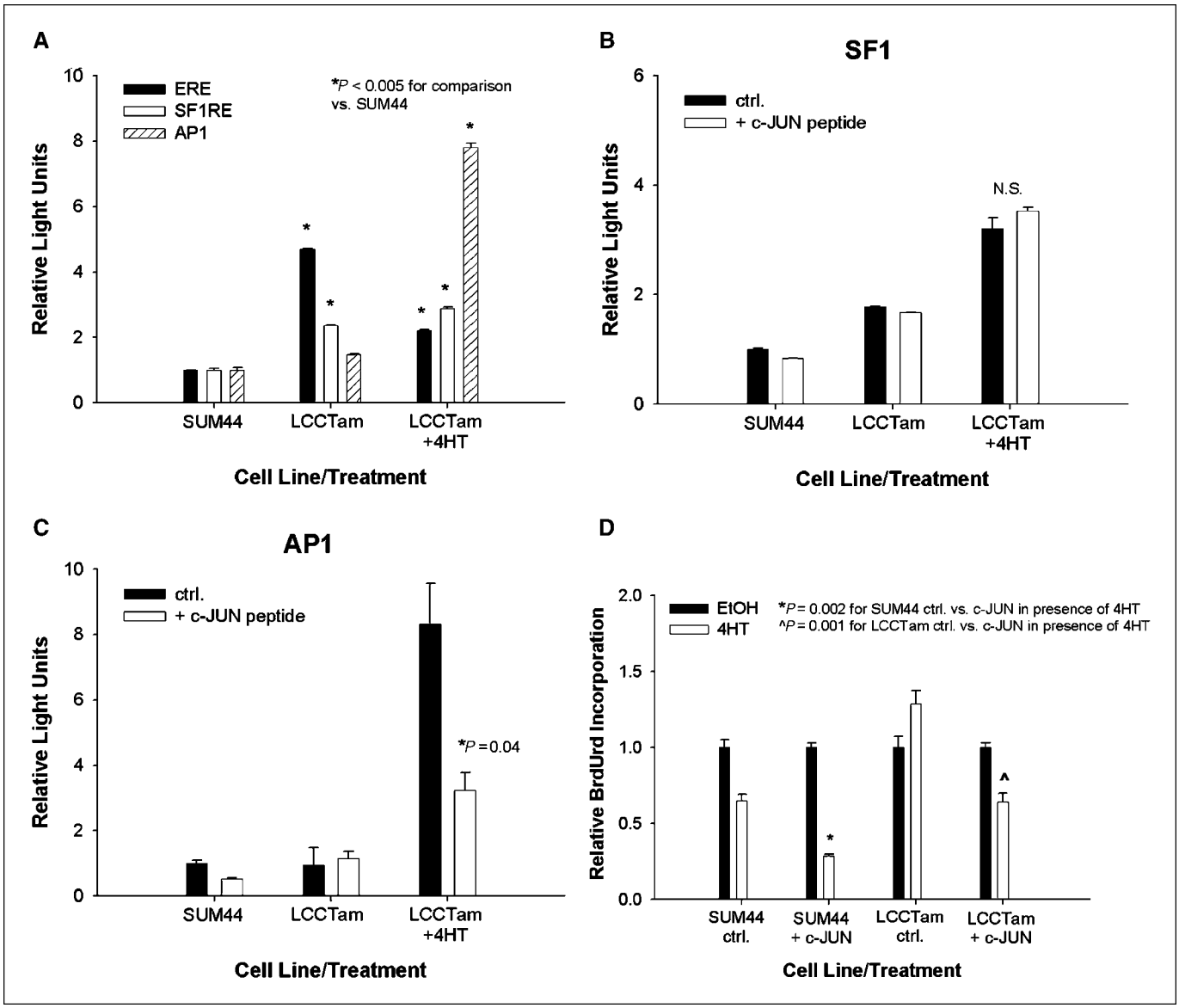


Figure 5. ERRγ-associated transcriptional activity is increased in resistant LCCTam cells, and inhibition of AP1 restores TAM sensitivity. *A,* cells were seeded and transfected with 3xERE-, 3xSF1RE-, and 3xAP1-luciferase and phRL-SV40 Renilla, incubated, lysed, and analyzed as described for Fig. 1*A*. *LCCTam+4HT,* cells that were cultured in SFIH containing 500 nmol/L 4HT. ANOVA $P < 0.001$, and $^*P < 0.005$ for all comparisons versus SUM44 by *post hoc* Student's *t* test. *B,* cells were seeded and transfected with 3xSF1RE-luciferase and phRL-SV40 Renilla for 1 d before treatment with 20 μmol/L c-JUN peptide or PBS control (*Ctrl.*). Cells were then incubated, lysed, and analyzed as described for Fig. 1*A*; LCCTam+4HT is as described above. *C,* cells were seeded and transfected with 3xAP1-luciferase, then treated, harvested, and analyzed as described. LCCTam+4HT is as described above. ANOVA $P < 0.001$; $^*P = 0.04$ for control versus c-JUN peptide by *post hoc* Student's *t* test. *D,* cells were seeded in 96-well dishes 24 h before treatment with 1,000 nmol/L 4HT or ethanol control in the presence of 20 μmol/L c-JUN peptide or PBS control. BrdUrd assays were performed as described above; *columns,* mean BrdUrd incorporation relative to ethanol for a representative experiment performed in quintuplicate; *bars,* SE. *, $P = 0.002$ for SUM44, control versus c-JUN in the presence of 4HT, and ^, $P = 0.001$ for LCCTam, control versus c-JUN in the presence of 4HT, by Student's *t* test.

A

HUGO symbol	Unigene ID	Description	LCCTam, fold change	AP1[a]
LRRC15	Hs.288467	leucine-rich repeat containing 15	26.557	4
LRP4	Hs.4930	low density lipoprotein receptor-related protein 4	13.311	3
LAMA4	Hs.213861	laminin, α4	9.891	3
HMGCS2	Hs.59889	3-Hydroxy-3-methylglutaryl-coenzyme A synthase 2	9.578	3
AMBP	Hs.436911	α-1-microglobulin/bikunin precursor	5.470	3
KRT13	Hs.463032	keratin 13	3.950	5
PTGES	Hs.146688	prostaglandin E synthase	3.238	2
CELSR2	Hs.57652	cadherin, EGF LAG seven-pass G-type receptor 2	2.696	2
TGFBR3	Hs.482390	transforming growth factor, βreceptor III	2.561	3
STARD8	Hs.95140	START domain containing 8	2.402	3
MSN	Hs.87752	moesin	2.385	2
FASN	Hs.83190	fatty acid synthase	2.110	2
CHST8	Hs.165724	carbohydrate (*N*-acetylgalactosamine 4-0) sulfotransferase 8	2.107	5
CTSA	Hs.517076	cathepsin A	2.104	3

[a]Number of AP1 response elements within 5,000 bp of the transcriptional start site.

B

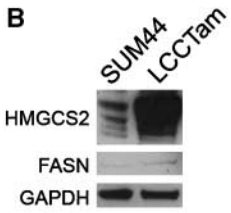


Figure 6. Genes overexpressed in LCCTam cells contain multiple AP1 response elements in their promoter regions. *A,* genes overexpressed by ≥2-fold in LCCTam versus SUM44 cells in the microarrays were screened for the presence of AP1 response elements in their proximal promoter regions using TRANSFAC Professional 11.1. *a,* the number of consensus sites found within 5,000 bp of the transcriptional start site. *B,* representative Western blot showing HMGCS2 and FASN expression, and the GAPDH loading control, in SUM44 and LCCTam cells.

hormone receptor coactivators and corepressors expressed by the tumor, differential metabolism of antiestrogens, and changes in the expression of additional genes that control cell proliferation and/or apoptosis (13). One or more of these mechanisms is likely contributing to the TAM-resistance phenotype of LCCTam cells. Relative to SUM44, the resistant LCCTam cells express 3-fold less ERα. However, SUM44 cells express high basal levels of ERα.[6] Consequently, the reduced level of ERα expression in LCCTam is comparable with that observed in MCF-7 breast cancer cells (~73% of basal MCF-7 ERα levels by qPCR; data not shown). Because ERα levels in MCF-7 cells are clearly sufficient to confer antiestrogen sensitivity, it is unlikely that ERα down-regulation in LCCTam is the major determinant of resistance in this model.

Our siRNA knockdown and cDNA overexpression studies are the first to show that ERRγ is an essential regulator of TAM responsiveness in lobular breast cancer cells. Until now, the role of ERRs (and specifically ERRγ) in breast cancer therapeutic response has not been well-understood. In 2002, Ariazi and colleagues (36) published a study of ERR family expression in 38 breast tumors compared with normal mammary epithelial cells (MEC). ERRγ mRNA expression is nearly 4-fold higher in breast tumors than in MECs and is positively associated with ER and PR expression. These authors conclude that the correlation of ERRγ with ER and PR is indicative of a better prognosis (36). Although this is certainly plausible, the presence of ER and PR do not always indicate hormone sensitivity in breast cancer. As discussed above, TAM therapy is ineffective in ~30% of patients with ER+/PR+ breast tumors, and the majority of initial responders who acquire resistance to TAM and other endocrine agents do so without losing detectable ER expression (10). 4HT-bound ERRγ is also known to activate transcription at AP1 sites, and elevated AP1 activity has previously been linked to TAM resistance *in vitro* (51, 52) and *in vivo* (53, 54). This is consistent with our findings that AP1 activity is robustly increased in the resistant LCCTam cells in the presence of 4HT, and that AP1 inhibition reverses the TAM-resistant phenotype of LCCTam cells while increasing the sensitivity of SUM44 cells to growth inhibition by this antiestrogen. To our knowledge, this is the first functional consequence of ERRγ-driven AP1 transcriptional activity that has been reported.

No endogenous ERRγ/AP1 target genes have yet been identified. The genes in Fig. 6*A* are strong candidates as ERRγ/AP1 targets in breast cancer. We confirmed the differential regulation of the endogenous HMGCS2 and FASN, and we propose that HMGCS2 and FASN are putative downstream targets of ERRγ in the resistant LCCTam cell line. Further assessment of their direct regulation by ERRγ/AP1 is in progress. HMGCS2 is a nuclear-encoded mitochondrial matrix gene that can regulate ketogenesis and cholesterol synthesis (45, 46) and FASN is the final enzyme of the fatty acid biosynthetic pathway (47). Components of all three processes have been implicated in the etiology or progression of breast cancer, and FASN activity can affect hormonal sensitivity in breast and endometrial cancer cells (48–50). Moreover, ERRγ has been shown to control the switch from fetal use of carbohydrates to lipid-dependent oxidative metabolism in the adult mouse heart by regulating a series of genes that drive fatty acid oxidation, oxidative phosphorylation, and mitochondrial electron transport (55). That ERRγ might also affect these metabolic processes in the context of breast cancer and TAM resistance is intriguing and will be the focus of future studies. Notably, this possibility is supported by a very recent publication by Montero and colleagues (56), which reports that increased mitochondrial cholesterol content promotes resistance to doxorubicin in hepatocellular carcinoma. Other genes in Fig. 6*A* also are of interest. High LRRC15 expression has been previously linked to invasive and aggressive behavior in breast and prostate cancer (57, 58), and MSN is a marker of basal-like breast cancers (59); our future studies will also pursue the role(s) of these genes in endocrine-resistant breast cancer and their regulation by ERRγ/AP1.

[6] D.A. Zajchowski and S.P. Ethier, unpublished data.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

Received 7/11/2008; accepted 8/20/2008.

Grant support: Ladies Auxiliary to the VFW, the Susan G. Komen Foundation (PDF0503551), and the Department of Defense Breast Cancer Research Program (BC051851; R.B. Riggins), and Public Health Service award CA096483-01A1 from the National Cancer Institute, and BC030280 from the Department of Defense Breast Cancer Research Program (R. Clarke). Technical services were provided by the Flow Cytometry, Macromolecular Analysis, Microscopy & Imaging, and Tissue Culture Shared Resources, which are supported by Public Health Service award 1P30-CA-51008-16 (Cancer Center Support Grant to the Lombardi Comprehensive Cancer Center).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

We thank past and current members of the laboratory for their critical comments and insightful discussions, as well as the staff of the Lombardi Comprehensive Cancer Center Flow Cytometry, Macromolecular Analysis, Microscopy and Imaging, and Tissue Culture Shared Resources for technical assistance.

References

1. Jemal A, Siegel R, Ward E, et al. Cancer statistics, 2008. CA Cancer J Clin 2008;58:71–96.

2. Li CI, Anderson BO, Daling JR, Moe RE. Trends in incidence rates of invasive lobular and ductal breast carcinoma. JAMA 2003;289:1421–4.

3. Biglia N, Mariani L, Sgro L, Mininanni P, Moggio G, Sismondi P. Increased incidence of lobular breast cancer in women treated with hormone replacement therapy: implications for diagnosis, surgical and medical treatment. Endocr Relat Cancer 2007;14:549–67.

4. Jimeno A, Amador ML, Gonzalez-Cortijo L, et al. Initially metastatic breast carcinoma has a distinct disease pattern but an equivalent outcome compared with recurrent metastatic breast carcinoma. Cancer 2004;100:1833–42.

5. Tubiana-Hulin M, Stevens D, Lasry S, et al. Response to neoadjuvant chemotherapy in lobular and ductal breast carcinomas: a retrospective study on 860 patients from one institution. Ann Oncol 2006;17: 1228–33.

6. Wenzel C, Bartsch R, Hussian D, et al. Invasive ductal carcinoma and invasive lobular carcinoma of breast differ in response following neoadjuvant therapy with epidoxorubicin and docetaxel + G-CSF. Breast Cancer Res Treat 2007;104:109–14.

7. Early Breast Cancer Trialists' Collaborative Group. Tamoxifen for early breast cancer. Cochrane Database Syst Rev 2001;CD000486.

8. Smith DB, Howell A, Wagstaff J. Infiltrating lobular carcinoma of the breast: response to endocrine therapy and survival. Eur J Cancer Clin Oncol 1987;23: 979–82.

9. Rakha EA, El-Sayed ME, Powe DG, et al. Invasive lobular carcinoma of the breast: response to hormonal therapy and outcomes. Eur J Cancer 2008;44:73–83.

10. Clarke R, Skaar TC, Bouker KB, et al. Molecular and pharmacological aspects of antiestrogen resistance. J Steroid Biochem Mol Biol 2001;76:71–84.

11. Clarke R, Liu MC, Bouker KB, et al. Antiestrogen resistance in breast cancer and the role of estrogen receptor signaling. Oncogene 2003;22:7316–39.

12. Riggins R, Bouton AH, Liu MC, Clarke R. Antiestrogens, aromatase inhibitors, and apoptosis in breast cancer. Vitam Horm 2005;71:201–37.

13. Riggins RB, Schrecengost RS, Guerrero MS, Bouton AH. Pathways to Tamoxifen resistance. Cancer Lett 2007; 256:1–24.

14. Lacroix M, Leclercq G. Relevance of breast cancer cell lines as models for breast tumours: an update. Breast Cancer Res Treat 2004;83:249–89.

15. Ethier SP, Mahacek ML, Gullick WJ, Frank TS, Weber BL. Differential isolation of normal liminal mammary epithelial cells and breast cancer cells from primary and metastatic sites using selective media. Cancer Res 1993; 53:627–35.

16. van de WM, Barker N, Harkes IC, et al. Mutant E-cadherin breast cancer cells do not display constitutive Wnt signaling. Cancer Res 2001;61:278–84.

17. Cowley SM, Parker MG. A comparison of transcriptional activation by ER α and ER β. J Steroid Biochem Mol Biol 1999;69:165–75.

18. Vanacker JM, Bonnelye E, Chopin-Delannoy S, Delmarre C, Cavailles V, Laudet V. Transcriptional activities of the orphan nuclear receptor ERR α (estrogen receptor-related receptor-α). Mol Endocrinol 1999;13:764–73.

19. Hong H, Yang L, Stallcup MR. Hormone-independent transcriptional activation and coactivator binding by novel orphan nuclear receptor ERR3. J Biol Chem 1999; 274:22618–26.

20. Riggins R, Zwart A, Nehra N, Agarwal P, Clarke R. The NFkB inhibitor parthenolide restores ICI 182,780 (Faslodex; Fulvestrant)-induced apoptosis in antiestrogen resistant breast cancer cells. Mol Cancer Ther 2005; 4:33–41.

21. Riggins RB, Thomas KS, Ta HQ, et al. Physical and functional interactions between Cas and c-Src induce tamoxifen resistance of breast cancer cells through pathways involving epidermal growth factor receptor and signal transducer and activator of transcription 5b. Cancer Res 2006;66:7007–15.

22. Vindelov LL, Christensen IJ, Nissen NI. A detergent-trypsin method for the preparation of nuclei for flow cytometric DNA analysis. Cytometry 1983;3:323–7.

23. Brünner N, Frandsen TL, Holst-Hansen C, et al. MCF7/LCC2: A 4-hydroxytamoxifen resistant human breast cancer variant which retains sensitivity to the steroidal antiestrogen ICI 182,780. Cancer Res 1993;53: 3229–32.

24. Figueiredo BC, Stratakis CA, Sandrini R, et al. Comparative genomic hybridization analysis of adrenocortical tumors of childhood. J Clin Endocrinol Metab 1999;84:1116–21.

25. Gomez BP, Riggins RB, Shajahan AN, et al. Human X-box binding protein-1 confers both estrogen independence and antiestrogen resistance in breast cancer cell lines. FASEB J 2007;21:4013–27.

26. Ellis M, Davis N, Coop A, et al. Development and validation of a method for using breast core needle biopsies for gene expression microarray analyses. Clin Cancer Res 2002;8:1155–66.

27. Liu A, Zhang Y, Gehan E, Clarke R. Block principal component analysis with application to gene microarray data classification. Stat Med 2002;21:3465–74.

28. Bouker KB, Skaar TC, Fernandez DR, et al. Interferon regulatory factor-1 mediates the proapoptotic but not cell cycle arrest effects of the steroidal antiestrogen ICI 182,780 (Faslodex, Fulvestrant). Cancer Res 2004;64: 4030–9.

29. Harris JR, Lippman ME, Morrow M, Osborne CK, editors. Diseases of the Breast. 2nd ed. Philadelphia: Lippincott Williams & Wilkins; 2000.

30. Forozan F, Veldman R, Ammerman CA, et al. Molecular cytogenetic analysis of 11 new breast cancer cell lines. Br J Cancer 1999;81:1328–34.

31. Brunner N, Boulay V, Fojo A, Freter CE, Lippman ME, Clarke R. Acquisition of hormone-independent growth in MCF-7 cells is accompanied by increased expression of estrogen-regulated genes but without detectable DNA amplifications. Cancer Res 1993;53:283–90.

32. Horard B, Vanacker JM. Estrogen receptor-related receptors: orphan receptors desperately seeking a ligand. J Mol Endocrinol 2003;31:349–57.

33. Giguère V. To ERR in the estrogen pathway. Trends Endocrinol Metab 2002;13:220–5.

34. Ariazi EA, Jordan VC. Estrogen-related receptors as emerging targets in cancer and metabolic disorders. Curr Top Med Chem 2006;6:203–15.

35. Ariazi EA, Kraus RJ, Farrell ML, Jordan VC, Mertz JE. Estrogen-related receptor {α}1 transcriptional activities are regulated in part via the ErbB2/HER2 signaling pathway. Mol Cancer Res 2007;5:71–85.

36. Ariazi EA, Clark GM, Mertz JE. Estrogen-related receptor α and estrogen-related receptor γ associate with unfavorable and favorable biomarkers, respectively, in human breast cancer. Cancer Res 2002;62: 6510–8.

37. Reiner GC, Katzenellenbogen BS. Characterization of estrogen and progesterone receptors and the dissociated regulation of growth and progesterone receptor stimulation by estrogen in MDA-MB-134 human breast cancer cells. Cancer Res 1986;46:1124–31.

38. Reis-Filho JS, Simpson PT, Turner NC, et al. FGFR1 emerges as a potential therapeutic target for lobular breast carcinomas. Clin Cancer Res 2006;12: 6652–62.

39. Greschik H, Wurtz JM, Sanglier S, et al. Structural and functional evidence for ligand-independent transcriptional activation by the estrogen-related receptor 3. Mol Cell 2002;9:303–13.

40. Greschik H, Flaig R, Renaud JP, Moras D. Structural basis for the deactivation of the estrogen-related receptor γ by diethylstilbestrol or 4-hydroxytamoxifen and determinants of selectivity. J Biol Chem 2004;279: 33639–46.

41. Holzberg D, Knight CG, Ttrich-Breiholz O, et al. Disruption of the c-JUN-JNK complex by a cell-permeable peptide containing the c-JUN δ domain induces apoptosis and affects a distinct set of interleukin-1-induced inflammatory genes. J Biol Chem 2003; 278:40213–23.

42. Huppunen J, Wohlfahrt G, Aarnisalo P. Requirements for transcriptional regulation by the orphan nuclear receptor ERRγ. Mol Cell Endocrinol 2004;219:151–60.

43. Matys V, Kel-Margoulis OV, Fricke E, et al. TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. Nucleic Acids Res 2006;34: D108–10.

44. Kel AE, Gossling E, Reuter I, Cheremushkin E, Kel-Margoulis OV, Wingender E. MATCH: a tool for searching transcription factor binding sites in DNA sequences. Nucleic Acids Res 2003;31:3576–9.

45. Mascaro C, Nadal A, Hegardt FG, Marrero PF, Haro D. Contribution of steroidogenic factor 1 to the regulation of cholesterol synthesis1. Biochem J 2000; 350 Pt 3:785–90.

46. Ortiz JA, Gil-Gomez G, Casaroli-Marano RP, Vilaro S, Hegardt FG, Haro D. Transfection of the ketogenic mitochondrial 3-hydroxy-3-methylglutaryl-coenzyme A synthase cDNA into Mev-1 cells corrects their auxotrophy for mevalonate. J Biol Chem 1994;269:28523–6.

47. Wakil SJ, Stoops JK, Joshi VC. Fatty acid synthesis and its regulation. Annu Rev Biochem 1983;52:537–79.

48. Kallinowskil F, Davel S, Vaupell P, Baessler KH, Wagner K. Glucose, lactate, and ketone body utilization by human mammary carcinomas *in vivo*. Adv Exp Med Biol 1985;191:763–73.

49. Duncan RE, El-Sohemy A, Archer MC. Dietary factors and the regulation of 3-hydroxy-3-methylglutaryl coenzyme A reductase: implications for breast cancer and development. Mol Nutr Food Res 2005;49:93–100.

50. Lupu R, Menendez JA. Targeting fatty acid synthase in breast and endometrial cancer: an alternative to

selective estrogen receptor modulators? Endocrinology 2006;147:4056–66.
51. Dumont JA, Bitonti AJ, Wallace CD, Baumann RJ, Cashman EA, Cross-Doersen DE. Progression of MCF-7 breast cancer cells to antiestrogen-resistant phenotype is accompanied by elevated levels of AP-1 DNA-binding activity. Cell Growth Differ 1996;7:351–9.
52. Zhou Y, Yau C, Gray JW, et al. Enhanced NF κ B and AP-1 transcriptional activity associated with antiestrogen resistant breast cancer. BMC Cancer 2007;7:59.
53. Johnston SR, Lu B, Scott GK, et al. Increased activator protein-1 DNA binding and c-Jun NH2-terminal kinase activity in human breast tumors with acquired tamoxifen resistance. Clin Cancer Res 1999;5:251–6.
54. Schiff R, Reddy P, Ahotupa M, et al. Oxidative stress and AP-1 activity in tamoxifen-resistant breast tumors *in vivo*. J Natl Cancer Inst 2000;92:1926–34.
55. Alaynick WA, Kondo RP, Xie W, et al. ERRγ directs and maintains the transition to oxidative metabolism in the postnatal heart. Cell Metab 2007;6:13–24.
56. Montero J, Morales A, Llacuna L, et al. Mitochondrial cholesterol contributes to chemotherapy resistance in hepatocellular carcinoma. Cancer Res 2008;68:5246–56.
57. Stanbrough M, Bubley GJ, Ross K, et al. Increased expression of genes converting adrenal androgens to testosterone in androgen-independent prostate cancer. Cancer Res 2006;66:2815–25.
58. Schuetz CS, Bonin M, Clare SE, et al. Progression-specific genes identified by expression profiling of matched ductal carcinomas *in situ* and invasive breast tumors, combining laser capture microdissection and oligonucleotide microarray analysis. Cancer Res 2006;66:5278–86.
59. Charafe-Jauffret E, Monville F, Bertucci F, et al. Moesin expression is a marker of basal breast carcinomas. Int J Cancer 2007;121:1779–85.

BIOINFORMATICS ORIGINAL PAPER Vol. 25 no. 4 2009, pages 526–532
doi:10.1093/bioinformatics/btn660

Data and text mining

Differential dependency network analysis to identify condition-specific topological changes in biological networks

Bai Zhang[1], Huai Li[2], Rebecca B. Riggins[3], Ming Zhan[2], Jianhua Xuan[1], Zhen Zhang[4], Eric P. Hoffman[5], Robert Clarke[3] and Yue Wang[1,*]

[1]Department of Electrical and Computer Engineering, Virginia Polytechnic Institute and State University, Arlington, VA 22203, [2]Bioinformatics Unit, RRB, National Institute on Aging, National Institutes of Health, Baltimore, MD 21224, [3]Lombardi Comprehensive Cancer Center and Department of Oncology, Physiology and Biophysics, Georgetown University, Washington, DC 20057, [4]Department of Pathology, Johns Hopkins Medical Institutions, Baltimore, MD 21231 and [5]Research Center for Genetic Medicine, Children's National Medical Center, Washington, DC 20010, USA

Received on November 3, 2008; revised on December 17, 2008; accepted on December 22, 2008

Advance Access publication December 26, 2008

Associate Editor: Jonathan Wren

ABSTRACT

Motivation: Significant efforts have been made to acquire data under different conditions and to construct static networks that can explain various gene regulation mechanisms. However, gene regulatory networks are dynamic and condition-specific; under different conditions, networks exhibit different regulation patterns accompanied by different transcriptional network topologies. Thus, an investigation on the topological changes in transcriptional networks can facilitate the understanding of cell development or provide novel insights into the pathophysiology of certain diseases, and help identify the key genetic players that could serve as biomarkers or drug targets.

Results: Here, we report a differential dependency network (DDN) analysis to detect statistically significant topological changes in the transcriptional networks between two biological conditions. We propose a local dependency model to represent the local structures of a network by a set of conditional probabilities. We develop an efficient learning algorithm to learn the local dependency model using the Lasso technique. A permutation test is subsequently performed to estimate the statistical significance of each learned local structure. In testing on a simulation dataset, the proposed algorithm accurately detected all the genes with network topological changes. The method was then applied to the estrogen-dependent T-47D estrogen receptor-positive (ER+) breast cancer cell line datasets and human and mouse embryonic stem cell datasets. In both experiments using real microarray datasets, the proposed method produced biologically meaningful results. We expect DDN to emerge as an important bioinformatics tool in transcriptional network analyses. While we focus specifically on transcriptional networks, the DDN method we introduce here is generally applicable to other biological networks with similar characteristics.

Availability: The DDN MATLAB toolbox and experiment data are available at http://www.cbil.ece.vt.edu/software.htm.

Contact: yuewang@vt.edu

Supplementary information: Supplementary data are available at *Bioinformatics* online.

*To whom correspondence should be addressed.

1 INTRODUCTION

Recent advances in high-throughput genomic technologies provide ample opportunities to study cellular activities at the individual gene expression and network levels, while also presenting new challenges for data analysis (Clarke *et al*., 2008). Discovering the mechanisms that orchestrate the activities of genes and proteins in cells remains one of the key goals of systems biology studies (Kitano, 2002). Several approaches have been proposed to model genetic regulatory networks (Li *et al*., 2008), such as Bayesian networks (Friedman, 2004; Friedman *et al*., 2000; Husmeier, 2003), probabilistic Boolean networks (Shmulevich *et al*., 2002), state–space models (Rangel *et al*., 2004) and network component analysis (Liao *et al*., 2003). These methods attempt to construct a static network that can explain various gene regulation programs.

However, genetic regulatory networks are context-specific and dynamic in nature (Beyer *et al*., 2007; Clarke *et al*., 2008). Under different conditions, different regulatory components and mechanisms are activated and the topology of the underlying gene regulatory network changes accordingly. For example, in response to diverse conditions in the yeast, transcription factors alter their interactions and rewire the signaling networks (Luscombe *et al*., 2004). While the inference of transcriptional networks using data from composite conditions could sometimes be contradictory due to changes in the underlying topology, most network learning algorithms assume an invariant network topology (Friedman *et al*., 2000; Rangel *et al*., 2004; Shmulevich *et al*., 2002). Therefore, some methods have been presented to learn condition-specific transcriptional networks in yeast (Kim *et al*., 2006; Segal *et al*., 2003). It is important to focus on and examine the topological changes in transcriptional networks between disease and normal conditions or under different stages of cell development. For example, a deviation from normal regulatory network topology may reveal the mechanism of pathogenesis (Hood *et al*., 2004), and the

genes that undergo the most network topological changes may serve as biomarkers or drug targets.

Several methods have been proposed to utilize network topology information to carry out various bioinformatics tasks. Liu *et al*. (2006) introduced a topology-based cancer classification method, where correlation networks were first constructed and later used to perform classification. Fuller *et al*. (2007) developed weighted gene co-expression network analysis strategies, via single network analysis and differential network analysis, to identify physiologically relevant modules. Qiu *et al*. (2005, 2007) proposed an ensemble dependence model to detect the dependence changes of gene clusters between cancer and normal conditions for cancer classification, and further extended the dependence model to dependence networks. Wei and Li (2007) introduced a Markov random field model for network-based analysis of genomic data that utilizes the known pathway structures to identify differentially expressed genes and sub-networks.

In this article, we propose a differential dependency network (DDN) analysis to model and detect the statistically significant topological changes in transcriptional networks between two conditions. We use local dependency models to characterize the dependencies of genes in the network and represent local network structures. Local dependency models decompose the whole network into a series of local networks, which serve as the basic elements of the network used for statistical testing. Unlike other dependency models that consider only pairwise relationships (Choi *et al*., 2005; Fuller *et al*., 2007; Kostka and Spang, 2004; Watson, 2006) or binding triples (Qiu *et al*., 2007), the local dependency models select the number of dependent variables automatically by the Lasso method (Tibshirani, 1996), and thereby learn the local network structures. Subsequently, we perform permutation tests on the local dependency models under two conditions and assign the *P*-values to the local structures. It may seem straightforward to construct an entire network under each condition and compare the differences between the two networks (Fuller *et al*., 2007; Qiu *et al*., 2007). However, in realistic applications this approach runs into the difficulty that the network structure learning can be inconsistent with a limited number of data samples. The detection procedure proposed here assures the statistical significance of the detected network topological changes by performing a permutation test on individual local structures. We also pinpoint 'hot spots' in the network where the genes exhibit network topological changes between two conditions above a given significance level. Lastly, we extract and visualize the DDN, i.e. the sub-network showing significant topological changes. We demonstrate the usefulness of the proposed method on both simulated and real microarray data. Tested on a simulation dataset, the proposed algorithm accurately captured the genes with network topological changes. When applied to the estrogen-dependent T-47D estrogen receptor-positive (ER+) breast cancer cell line datasets and human and mouse embryonic stem cell (ESC) datasets, the DDN analysis obtained biological meaningful and promising results.

2 METHODS

2.1 Local dependency models

Given a set of random variables $\mathbf{X}=\{X_1,X_2,\ldots,X_M\}$, a dependency network for X is modeled by a set of local conditional probability distributions, one for each node given its parents, denoted as $\mathbf{Z}_i$, which satisfies

$$P\left(X_i|\mathbf{Z}_i\right)=P\left(X_i|\mathbf{X}_{-i}\right) \tag{1}$$

where $\mathbf{X}_{-i}=\{X_1,X_2,\ldots,X_{i-1},X_{i+1},\ldots,X_M\}$ and $\mathbf{Z}_i\subseteq\mathbf{X}_{-i}$. $P(X_i|\mathbf{Z}_i)$ also represents the local structure of node X_i, i.e. the relationship of node X_i and its parents $\mathbf{Z}_i$ on the graph (Heckerman *et al*., 2000).

Inspired by this formulation, we propose a local dependency model to describe the dependencies of genes in a transcriptional network. Unlike a conventional dependency network approach, where there is only one conditional probability distribution for each node given its parents, our local dependency model allows more than one conditional probability distributions for each node. Mathematically, suppose there are *M* genes in the network of interest, and the dependencies of gene *i* on other genes are formulated by a set of conditional probabilities,

$$\boldsymbol{\mathcal{P}}_i=\left\{P\left(X_i|\mathbf{Z}_{i,1}\right),P\left(X_i|\mathbf{Z}_{i,2}\right),\ldots,P\left(X_i|\mathbf{Z}_{i,s_i}\right)\right\},\ i=1,2,\ldots,M \tag{2}$$

where $\mathbf{Z}_{i,1},\mathbf{Z}_{i,2},\ldots,\mathbf{Z}_{i,s_i}$ are some subsets of $\mathbf{X}_{-i}$ and s_i is the number of conditional probabilities for random variable X_i. We use X_i to refer both to the expressions of gene *i* and to its corresponding node on the graph. This modification is primarily based on the following considerations. First, our goal is not to construct the entire network that represents the full joint distribution of all variables, rather we wish to model the local structures for further statistical testing. Second, many genes are highly correlated and the data points are very limited when extracting most biological networks. Through our experiments, we found that the conventional approach misses some meaningful dependency connections in data-sparse situations. For example, regulator genes R1 and R2 have the same target gene A, and the expression patterns of R1, R2 and A are highly correlated. When the data points are few, the standard approach may only select one of the dependencies, for instance, gene A on gene R1, even though the dependency of gene A on gene R2 is only slightly less significant than the dependency of gene A on gene R1. However, the dependencies of gene A on genes R1 and R2 are both important, and we want to keep the rich structural information for later step to assess the topological changes. Therefore, to retain more meaningful local structure information, instead of selecting 'the best' local structure, we select a set of 'sufficiently good' local structures for further statistical testing. We achieve this goal by allowing each node to be modeled by more than one conditional probability distribution.

2.2 Local structure learning

The conditional probability distributions in Equation (2) can be inferred by regression methods. In our approach, we consider a linear regression model in which the variable X_i is predicted by a linear function of $\mathbf{Z}_i$

$$X_i=\boldsymbol{\beta}^{\mathbf{T}}\mathbf{Z}_i+\varepsilon_i,\quad i=1,2,\ldots,M \tag{3}$$

where $\mathbf{Z}_i\in\{\mathbf{Z}_{i,1},\mathbf{Z}_{i,2},\ldots,\mathbf{Z}_{i,s_i}\}$ is a column vector of random variables, $\boldsymbol{\beta}$ is a column vector of unknown parameters. The random error ε_i is independent of $\mathbf{Z}_i$ and is assumed to have normal distribution $N(0,\sigma_i^2)$. The local conditional probability $P(X_i|\mathbf{Z}_i)$ is given by

$$P\left(X_i|\mathbf{Z}_i\right)=N\left(\boldsymbol{\beta}^{\mathbf{T}}\mathbf{Z}_i,\sigma_i^2\right) \tag{4}$$

Learning the structure of the local dependency model requires the selection of a $\mathbf{Z}_i$ that shows good predictability of X_i. Given a predefined maximum size of $\mathbf{Z}_i$, *K*, we examine all C_{M-1}^K combinations of the elements in $\mathbf{X}_{-i}$ with size *K*. *K* can be empirically set to a positive integer between 1 and $M-1$. When $K=1$, the proposed local dependency model only considers pairwise relationships. When $K=M-1$, the proposed local dependency model is equivalent to standard dependency networks as described in Equation (1) (Heckerman *et al*., 2000).

Suppose one *K*-combination of $\mathbf{X}_{-i}$ is $\{X_{k_1},X_{k_2},\ldots,X_{k_K}\}$, where $k_1,k_2,\ldots,k_K\in\{1,2,\ldots,i-1,i+1,\ldots,M\}$, and there are *N* expression samples. Lower case letter $x_i(j)$ denotes the *j*-th sample value taken by the

variable X_i, $j=1,2,\ldots,N$. We perform a L_1 constrained regression of X_i on $\mathbf{Z}_i=\{X_{k_1},X_{k_2},\ldots,X_{k_K}\}$

$$\hat{\boldsymbol{\beta}}_{\text{Lasso}}=\text{argmin}\left\{\sum_{j=1}^{N}\left(x_i(j)-\sum_{l=1}^{K}\beta_l x_{k_l}(j)\right)^2\right\},\ s.t.\ \sum_{l=1}^{K}|\beta_l|\leqslant t. \quad (5)$$

Equation (5) is known as the Lasso estimator (Tibshirani, 1996), which minimizes L_2 norm loss with constraint on the L_1 norm of $\boldsymbol{\beta}$. The nature of L_1 constraint tends to make some coefficients in $\hat{\boldsymbol{\beta}}_{\text{Lasso}}$ exactly zero, and hence it automatically selects a subset of features and leads to a simpler model that avoids overfitting the data, and therefore usually has better generalization performance. The parameter $t\geq 0$ controls the amount of shrinkage that is applied to the estimates. In our software implementation, parameter t is determined by 5-fold cross-validation. Solving Equation (5) is a convex optimization problem, and can be solved very efficiently. We adopt the least angle regression (LARS) method to solve this problem. The detailed procedure of LARS can be found in Efron *et al*. (2004).

We also use a prescreening strategy to release the computational burden. We first regress X_i on $\mathbf{Z}_i=\{X_{k_1},X_{k_2},\ldots,X_{k_K}\}$, using the ordinary least square method

$$\hat{\boldsymbol{\beta}}_{\text{OLS}}=\text{argmin}\left\{\sum_{j=1}^{N}\left(x_i(j)-\sum_{l=1,2,\ldots,K}\beta_l x_{k_l}(j)\right)^2\right\}. \quad (6)$$

If the corresponding mean square error (MSE) is above a predetermined threshold T, which means X_i cannot be accurately predicted by the subset $\{X_{k_1},X_{k_2},\ldots,X_{k_K}\}$, the subset $\{X_{k_1},X_{k_2},\ldots,X_{k_K}\}$ will be discarded. If the MSE is below T, we will then perform the L_1 constrained regression of X_i.

We perform the above prescreening and local structure learning with the Lasso on each of K-combinations of $\mathbf{X}_{-i}$, and obtain predictor sets $\mathbf{Z}_{i,1},\mathbf{Z}_{i,2},\ldots,\mathbf{Z}_{i,s_i}$ and the conditional probability distributions $\boldsymbol{\mathcal{P}}_i=\{P(X_i|\mathbf{Z}_{i,1}),P(X_i|\mathbf{Z}_{i,2}),\ldots,P(X_i|\mathbf{Z}_{i,s_i})\}$ for node X_i.

To measure how well variables $\mathbf{Z}_i$ can predict X_i, or how well the local dependency model fits gene expression microarray data, we further introduce the definition of coefficient of determination (COD)

$$\text{COD}=\frac{var\{X_i\}-var\left\{X_i-f_{X_i|\mathbf{Z}_i}(\mathbf{Z}_i)\right\}}{var\{X_i\}} \quad (7)$$

where $var\{\cdot\}$ is the estimate of the variance of the random variable and $f_{X_i|\mathbf{Z}_i}(\cdot)$ is the best function in a given function class that minimizes the residual variance. COD has been successfully used in non-linear signal processing and probabilistic Boolean network inference (Shmulevich *et al*., 2002). Here we only use linear functions, and $\text{var}\{X_i-f_{X_i|\mathbf{Z}_i}(\mathbf{Z}_i)\}$ is an estimate of σ_i^2 in Equation (4).

2.3 Detection of statistically significant topological changes

To detect the statistically significant network topological changes between two experimental conditions, we assume there are M genes in the network of interest, and N_1 samples from condition 1 and N_2 samples from condition 2. We further denote the datasets from two conditions by $\mathbf{D}^{(m)}=[\mathbf{x}^{(m)}(1),\mathbf{x}^{(m)}(2),\ldots,\mathbf{x}^{(m)}(N_m)]$, where superscript (m) indicates condition m, m= 1, 2. The bold face lower case letter $\mathbf{x}^{(m)}(j)$ denotes the column vector $[x_1^{(m)}(j),x_2^{(m)}(j),\ldots,x_M^{(m)}(j)]^T$, where lower case letter $x_i^{(m)}(j)$ denotes the j-th sample value taken by variable X_i under condition m.

By applying the learning procedure to datasets $\mathbf{D}^{(1)}$ and $\mathbf{D}^{(2)}$, respectively, we obtain $\boldsymbol{\mathcal{P}}_i^{(1)}=\{P(X_i|\mathbf{Z}_{i,1}^{(1)}),P(X_i|\mathbf{Z}_{i,2}^{(1)}),\ldots,P(X_i|\mathbf{Z}_{i,s_i^{(1)}}^{(1)})\}$ under condition 1 and $\boldsymbol{\mathcal{P}}_i^{(2)}=\{P(X_i|\mathbf{Z}_{i,1}^{(2)}),P(X_i|\mathbf{Z}_{i,2}^{(2)}),\ldots,P(X_i|\mathbf{Z}_{i,s_i^{(2)}}^{(2)})\}$ under condition 2 for each node i, $i=1,2,\ldots,M$. Then we take the union of the local structures learned under two conditions

$$\boldsymbol{\mathcal{P}}_i=\boldsymbol{\mathcal{P}}_i^{(1)}\cup\boldsymbol{\mathcal{P}}_i^{(2)},\ i=1,2,\ldots,M, \quad (8)$$

for further statistical testing.

For each conditional probability distribution in $\boldsymbol{\mathcal{P}}_i$, $i=1,2,\ldots,M$, for instance, $P(X_i|\mathbf{Z}_i)\in\boldsymbol{\mathcal{P}}_i$, we perform a permutation test to assess how significantly it is different between two conditions. Given samples $\{[x_i^{(1)}(j^{(1)}),\mathbf{z}_i^{(1)}(j^{(1)})]^T,\ j^{(1)}=1,2,\ldots,N_1\}$ under the first condition and $\{[x_i^{(2)}(j^{(2)}),\mathbf{z}_i^{(2)}(j^{(2)})]^T,\ j^{(2)}=1,2,\ldots,N_2\}$ under the second condition, we calculate $\text{COD}^{(1)}$ and $\text{COD}^{(2)}$, using Equation (7). A test statistic $\hat{\theta}$ is defined by the absolute difference of the coefficients of determination under two conditions

$$\hat{\theta}=\left|\text{COD}^{(1)}-\text{COD}^{(2)}\right| \quad (9)$$

We want to test the null hypothesis, H_0, of no difference between $P^{(1)}(X_i|\mathbf{Z}_i)$ and $P^{(2)}(X_i|\mathbf{Z}_i)$. We first combine $\{[x_i^{(1)}(j^{(1)}),\mathbf{z}_i^{(1)}(j^{(1)})]^T,\ j^{(1)}=1,2,\ldots,N_1\}$ and $\{[x_i^{(2)}(j^{(2)}),\mathbf{z}_i^{(2)}(j^{(2)})]^T,j^{(2)}=1,2,\ldots,N_2\}$, and then randomly permute samples from two conditions and divide the data into two sets of N_1 and N_2 samples, respectively. We perform the above procedure B times, where B is set to 5000 in our software implementation, and calculate $\hat{\theta}_b^*$, $b=1,2,\ldots,B$ according to Equation (9). An estimate of the achieved significance level (ASL) of the test is

$$\text{ASL}=\frac{\sum_{b=1}^{B}\mathbf{1}_{\{\hat{\theta}_b^*\geq\hat{\theta}\}}}{B} \quad (10)$$

where the random variable $\hat{\theta}_b^*$ is generated by permutation and $\mathbf{1}_{\{\hat{\theta}_b^*\geq\hat{\theta}\}}$ denotes the indicator function, which takes 1 when $\hat{\theta}_b^*\geq\hat{\theta}$ and 0 otherwise. The smaller the value of ASL, the stronger the evidence against H_0 is. Equation (10) also is an estimate of the P-value. The detailed permutation procedure is described in Efron and Tibshirani (1993). This detection procedure is performed on every local structure in $\boldsymbol{\mathcal{P}}_i$, $i=1,2,\ldots,M$, and each local structure is assigned a P-value.

2.4 Identification of the 'hot spots' in the network and extraction of the DDN

Given a user-defined P-value cutoff, we obtain a set of statistically significant differential local structures. The nodes in these differential local structures are identified as 'hot spots' in the network, which are the genes undergoing topological changes defined by a specified significance level. These genes may correspond to the genes in disease- or process-related pathways.

DDN is the focused sub-network that exhibits the topological changes. We consider a connection to exist from each element in $\mathbf{Z}_i$ to X_i under one specific condition if the variance of $P(X_i|\mathbf{Z}_i)$ is below the user-defined threshold T for that condition (see Supplementary Material for discussions on the selection of T). We use different colors to represent connections appearing under different conditions. DDN provides a way to visualize the topological changes, and when applied to disease studies, DDN extracts and focuses on the disease-related pathways that may contribute to the understanding of the mechanism of the disease.

3 RESULTS

3.1 A simulation experiment

We first used the software SynTReN (Van den Bulcke *et al*., 2006) to generate one simulation dataset of a sub-network drawn from an existing signaling network in *Saccharomyces cerevisiae*. Then we changed part of network topology and used SynTReN to generate another dataset according to this modified network. The network topology under two conditions is shown in Figure 1. The network contains 20 nodes that represent 20 genes. The black lines indicate the regulatory relationships that exist under both conditions. The red and green lines are the regulatory relationships that only exist under conditions 1 and 2, respectively. The sub-network comprised of nodes MBP1_SWI6, CLB5, CLB6, PHO2, FLO1, FLO10 and

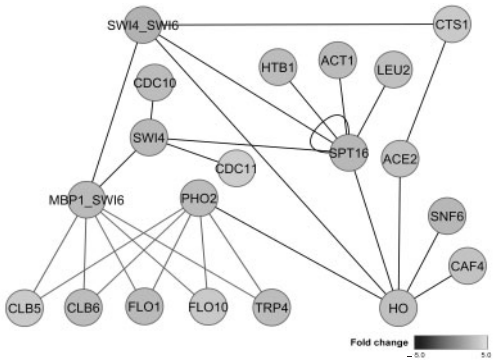


Fig. 1. The network topology under two conditions in the simulation study. Nodes in the network represent genes. Lines in the network indicate regulatory relationships between genes. The black lines are the regulatory relationships that exist under both conditions. The red and green lines represent the regulatory relationships that only exist under conditions 1 and 2, respectively. The DDN between the two conditions is the sub-network comprised of nodes MBP1_SWI6, CLB5, CLB6, PHO2, FLO1, FLO10 and TRP4 and green and red lines.

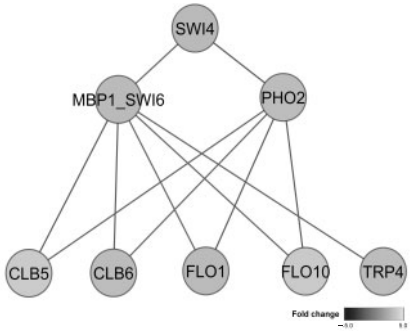


Fig. 2. The DDN extracted by the proposed algorithm in the simulation study. The red lines represent the connections (dependencies) that only exist under condition 1, and the green lines represent the connections (dependencies) that only exist under condition 2. The proposed DDN analysis successfully detected 9 of 10 connections that are different between two conditions and all the genes involved in the network topology changes. The connections between PHO2 and SWI4 under condition 1 (red) and between MBP_SWI6 and SWI4 under condition 2 (green) were falsely detected and the connection between PHO2 and TRP4 under condition 1 (red) was falsely missed.

Table 1. 'Hot spots' identified by DDN analysis in simulation study

Gene	Fold change	*P*-value (*t*-test)	Gene	Fold change	*P*-value (*t*-test)
CLB5	8.92E-01	6.92E-043	MBP1_SWI6	3.74E-02	4.58E-001
CLB6	4.79E-02	2.00E-001	PHO2	1.54E-01	1.34E-005
FLO1	−5.04E-02	4.44E-001	SWI4	−1.23E-01	3.64E-007
FLO10	7.73E-01	3.52E-024	TRP4	8.26E-02	1.00E-002

TRP4 and green and red lines is the DDN that our algorithm tries to identify from expression data.

The parameters for our algorithm are: threshold T is 0.25, P-value cutoff is 0.01 and the maximum size of $\mathbf{Z}_i$, K, is 2. Table 1 presents the 'hot spots' identified by the DDN analysis. Table 1 also shows the fold-changes of individual genes (after base 2 logarithm), and P-values of t-tests of individual genes. Our algorithm picked up all genes involved in topological changes, including some genes that did not show a significant difference in fold-change or t-tests, such as CLB6, FLO1 and MBP1_SWI6. This indicates that our algorithm can successfully detect these interesting genes using their topological information, even though the means of their expressions did not change substantially between the two conditions. Therefore, this method is able to identify biomarkers that cannot be picked up by traditional gene ranking methods, providing a complimentary approach for biomarker identification problem.

Figure 2 shows the DDN between the two conditions extracted by the proposed algorithm. The DDN shows network topological changes and the genes involved therein. The red lines in Figure 2 represent the connections that exist only under condition 1, and the green lines represent the connections that exist only under condition 2. Compared with the known network topology shown in Figure 1, the proposed algorithm correctly identified and extracted all the nodes with topology changes and 9 of 10 differential connections, with only the connection between PHO2 and TRP4 under condition 1 falsely missed, and the connection between PHO2 and SWI4 under condition 1 and the connection between MBP1_SWI6 and SWI4 under condition 2 falsely detected.

3.2 Breast cancer dataset analysis

We applied our method to the dataset from an ER+ breast cancer cell study by Lin *et al*. (2004). In that dataset, the estrogen-dependent T-47D ER+ breast cancer cell line was treated with 17β-estradiol (E2) and with E2 in combination with the pure anti-estrogen ICI 182 780 (ICI, Faslodex, Fulvestrant). Samples were then harvested on an hourly basis for the first 8 h (0–8 h) and bi-hourly for the next 16 h (10–24 h) for a total of 16 time points under each condition. Experiments were performed on microarrays generated by spotting the Compugen 19 K human oligo library, made by Sigma-Genosys, on poly-L-lysine-coated glass slides.

Here we are interested in the cellular response to the drug ICI, which inhibits E2 signaling through the ER (Howell, 2006). We first selected 55 genes that are reported in the literature to be relevant to breast cancer and responsiveness to ICI (for example, Kuo, 2007; Riggins *et al*., 2005, 2007). We then applied our DDN analysis to the data under two conditions (E2 versus E2+ICI). The parameters in our algorithm are: threshold T is 0.25, P-value cutoff is 0.01 and K is 2.

Table 2 lists the genes that exhibit significant topological changes in the network identified by DDN analysis. The DDN under these two conditions is shown in Figure 3. The genes identified by the proposed algorithm and their expression results (Table 2) are consistent with published data. For example, XBP1 and BCL2 show strongly decreased expression in response to E2+ICI relative to E2 alone, and both of these genes are known to be induced by E2 (Gompel *et al*., 2000; Tozlu *et al*., 2006; Wang *et al*., 2004).

In Figure 3, there are 18 red connections in the DDN, which implies that these connections exist only under E2 condition and disappear after the addition of ICI. Since ICI 182 780 is an ER antagonist, which works both by downregulating and by degrading the ER-alpha, it is plausible that these connections disappear because ICI is blocking or inactivating their connections. For example, as

Table 2. 'Hot spots' identified by DDN analysis in breast cancer study (see Supplementary Material for gene annotations)

Gene	Fold change	*P*-value (*t*-test)	Gene	Fold change	*P*-value (*t*-test)
ABCB11	2.05E-01	8.48E-01	ESR2	−3.72E-01	8.74E-01
AKT1	−4.02E+00	3.92E-01	F12	4.10E+00	3.81E-01
BCAR3	−4.92E-01	9.37E-01	HOXA10	4.14E+00	7.89E-01
BCL2	−2.46E+00	2.62E-01	MAPK1	−1.35E+00	6.09E-01
BIK	−2.75E+00	7.52E-01	MAPK13	2.81E+00	5.01E-01
BIRC1	5.98E-01	8.67E-01	MAPK14	−1.12E-01	9.35E-01
BIRC3	2.66E+00	5.60E-01	MAPK3	2.42E-01	9.67E-01
CAV3	4.12E+00	7.92E-01	MAPK4	1.55E+00	4.84E-01
CGA	4.19E+00	7.00E-01	MAPK8	−6.73E+00	1.65E-01
COX7A2L	3.94E+00	2.32E-01	NFKB1	3.91E-02	9.88E-01
EBAG9	2.04E+00	6.76E-01	NFKB2	−9.15E-02	6.92E-01

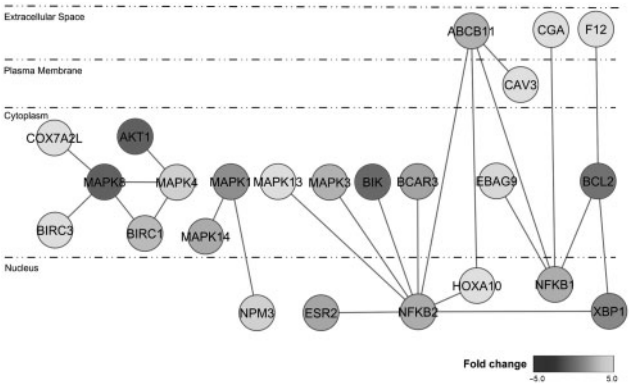


Fig. 3. DDN between breast cancer cell line treated with E2 and cell line treated with E2+ICI. The red lines represent the connections that exist only in breast cancer cell line treated with E2, and the green lines represent the connections that exist only in breast cancer cell line treated with E2+ICI.

a transcription factor, XBP1 can directly regulate gene expression through binding to its response element (Iwakoshi *et al*., 2003), or it can act as a co-regulator of other transcription factors, most notably ER-alpha, to enhance their transcriptional activity (Ding *et al*., 2003; Fang *et al*., 2004). Because BCL2 contains response elements for both ER-alpha and XBP1 (Gomez *et al*., 2007; Somai *et al*., 2003), the connection between XBP1 and BCL2 in the DDN may either be direct or involve ER-alpha as a latent variable, or intervening gene. In direct support of this predicted edge, we have shown that constitutive overexpression of XBP1 in a different breast cancer cell line (MCF-7) led to significantly increased mRNA and protein expression of both ER-alpha and BCL2 and functionally conferred antiestrogen resistance and estrogen-independence (Gomez *et al*., 2007).

Novel relationships between these genes identified by our DDN analysis will also serve as useful guidance for future studies. For example, BCAR3 is a well-established effector of cell motility, estrogen independence and antiestrogen resistance in ER+ breast cancer cell lines (Felekkis *et al*., 2005; Riggins *et al*., 2003; Schrecengost *et al*., 2007; Van Agthoven *et al*., 2006). Expression of NFKB2 and its activator BCL3 are also associated with estrogen independence in breast cancer cell lines (Pratt *et al*., 2003), and these nuclear factor κ B subunits appear to be selectively activated in clinical breast cancer (Pratt *et al*., 2003). However, there is no experimental evidence linking BCAR3 with NFKB2, so the suggestion that these two genes exhibit differential dependence under E2-treated conditions (Fig. 3) provides a starting point for biological studies of their relationship.

Additional relationships that may be completely new to breast cancer are also identified by this method. For example, MAPK8 (also known as JNK1) has been shown to be activated by BIRC1 (also known as NAIP) during its inhibition of caspase-mediated cell death (Sanna *et al*., 2002). In chronic fatigue syndrome, growth factor receptor signaling can activate MAPK4, which via Ras and/or PI3K can subsequently increase AKT1 activity (Englebienne and Meirleir, 2002). And finally, in B cells from patients with chronic lymphocytic leukemia NFKB1 (p50) homodimers are able to stimulate transcription from the BCL2 promoter through binding to another member of the BCL family (BCL3) (Viatour *et al*., 2003).

3.3 Human and mouse ESC analysis

ESCs can either maintain their pluripotency by self-renewal or undergo differentiation. The molecular mechanisms controlling ESC self-renewal and differentiation are complex and poorly understood (Sun *et al*., 2006; Zhan, 2008). ESCs harvested from different species show common characteristics, yet significant differences exist. Thus, cross-species analysis may help to distinguish between fundamental and species-specific mechanisms regulating ESC development (Sun *et al*., 2007; Zhan *et al*., 2005). Network biology can provide a new avenue for exploring ESC biology (Barabasi and Oltvai, 2004). Here, we used our new algorithm to conduct a human–mouse comparative analysis of ESCs, identifying evolutionarily divergent sub-networks. We focused our analysis on the cell cycle, a critical process for controlling cell development. In this study, 58 cell-cycle genes were selected for the DDN analysis. The 58 genes are the core components of the cell cycle machinery, and are orthologous between human and mouse cells. The expression profile data for these genes were determined from 18 samples from human ESCs and their earliest differentiation counterparts, embryoid bodies (EBs) and 18 samples from mouse ESCs and EBs, so that our inferred networks were directly related to ESC differentiation. The human ESC and EB expression data were determined from BG01, BG02 and BG03 cell lines in our previous studies using Illumina's BeadArrays (Liu *et al*., 2006), and from H1 (Sato *et al*., 2003) and HES2 (E-MEXP-303 of the ArrayExpress database) cell lines using Affymetrix chips. The mouse ESC and EB expression data were determined from V6.5 (GSE3231 of GEO database), R1 (GSE2972) and J1 (GSE3749) cell lines, based on Affymetrix chips. The final datasets contained 9 ESC and 9 EB (14-day differentiated) samples from human and mouse cells, respectively. In the network analysis, we set K to 1, and threshold T to 0.2 and P-value cutoff to 0.01.

Figure 4 shows DDN of the cell cycle between human and mouse cells (see Supplementary Material for gene annotations). The red lines represent the gene connections in human, and the green lines represent the connections in mouse. As shown, CDC25C, DUSP1 and BUB1 exhibit high connectivity on the network of human cells. On the other hand, PLK1, CDK2AP1, CDC20, TFDP1 and CDC5L showed a high connectivity on the network of mouse cells. These results suggest evolutionary divergence across species during ESC development and may provide clues for insights into species-specific

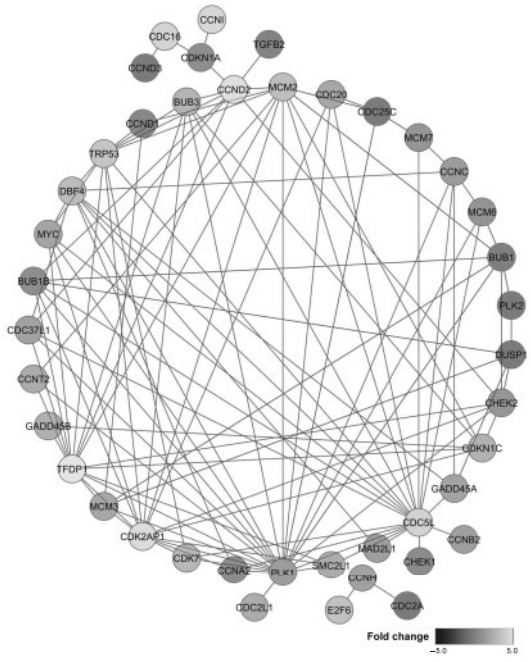


Fig. 4. DDN between human and mouse ES/EB cells. The red lines represent the connections that exist only in human ES/EB cells, and the green lines represent the connections that exist only in mouse ES/EB cells.

mechanism of the cell cycle in controlling ESC self-renewal and differentiation.

4 DISCUSSIONS

In this article, we propose a systematic approach to detect the statistically significant changes in transcriptional networks between two different experimental conditions. We tested our algorithm on simulation data, breast cancer data and ESC data. From the simulation study, we see that the proposed algorithm can capture the topological changes efficiently and accurately, even when the fold change of the expression values of each gene is not statistically significant. This approach utilizes the network structure information and provides an alternative way for biomarker identification. In addition, as knowledge of cellular networks accumulates, many biological databases will expand to contain more useful information. The proposed approach is an open framework, into which biological knowledge in specific applications can be easily incorporated as the local structure learning constraints.

The high level of correlation among genes is a common feature of microarray data. Therefore, we propose a local dependency model that allows multiple predictor sets for each node. Accordingly, a local structure learning algorithm is also represented. Lasso is used to select features for the predictor sets (Tibshirani, 1996), an approach that has been successfully applied to variable selection and graph structure learning (Meinshausen and Buhlmann, 2006). In the linear Gaussian case, under certain conditions, it is proved that the probability of estimating the correct neighborhood converges exponentially to 1, and as a consequence it is possible to obtain a consistent estimation of the full edge set (Meinshausen and Buhlmann, 2006). However, in microarray data, the so-called irrepresentable condition (Zhao and Yu, 2006) or the neighborhood stability assumption (Meinshausen and Buhlmann, 2006) can easily be violated in the presence of highly correlated genes. Some modified algorithms have been proposed to deal with the highly correlated cases, for example, elastic net (Zou and Hastie, 2005) and network-constrained regularization (Li and Li, 2008), both of which tend to group highly correlated predictors in the regression process. However, these two approaches are not suitable for our problem, because the grouping of highly correlated variables can be different under two conditions and this makes the later statistical testing problematic. The local structure learning algorithm proposed here attempts to alleviate the effects of the highly correlated gene expression data and to preserve local structure information for further statistical testing.

Some issues are worth further exploration. In this article, only linear relationships are considered. How non-linear relationships should be modeled efficiently and correctly, remains a difficult problem. Second, since many cellular reactions take place in the genome, transcriptome and proteome, it is essential to construct pathways by integrating data from heterogeneous sources.

In sum, this article presents a new approach to extract knowledge of a biological network by emphasizing the dynamic nature of cellular networks and utilizing a network's structural information. It also provides an alternative and promising approach to identify possible biomarkers and drug targets.

Funding: National Institutes of Health (CA109872, EB000830, CA096483, CA86323 and NS29525, partial); Department of Defense Breast Cancer Research Program BC030280. IRP/NIA/NIH (to H.L. and M.Z.).

Conflict of Interest: none declared.

REFERENCES

Barabasi,A.L. and Oltvai,Z.N. (2004) Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.*, **5**, 101–113.

Beyer,A. *et al.* (2007) Integrating physical and genetic maps: from genomes to interaction networks. *Nat. Rev. Genet.*, **8**, 699–710.

Choi,J.K. *et al.* (2005) Differential coexpression analysis using microarray data and its application to human cancer. *Bioinformatics*, **21**, 4348–4355.

Clarke,R. *et al.* (2008) The properties of high-dimensional data spaces: implications for exploring gene and protein expression data. *Nat. Rev. Cancer*, **8**, 37–49.

Ding,L.H. *et al.* (2003) Ligand-independent activation of estrogen receptor alpha by XBP-1. *Nucleic Acids Res.*, **31**, 5266–5274.

Efron,B. and Tibshirani,R. (1993) *An Introduction to the Bootstrap*. Chapman & Hall, New York.

Efron,B. *et al.* (2004) Least angle regression. *Ann. Stat.*, **32**, 407–451.

Englebienne,P. and Meirleir,K. (2002) *Chronic Fatigue Syndrome: A Biological Approach*. CRC Press, Boca Raton.

Fang,Y. *et al.* (2004) XBP-1 increases ER alpha transcriptional activity through regulation of large-scale chromatin unfolding. *Biochem. Biophys. Res. Commun.*, **323**, 269–274.

Felekkis,K.N. *et al.* (2005) AND-34 activates phosphatidylinositol 3-kinase and induces anti-estrogen resistance in a SH2 and GDP exchange factor-like domain-dependent manner. *Mol. Cancer Res.*, **3**, 32–41.

Friedman,N. (2004) Inferring cellular networks using probabilistic graphical models. *Science*, **303**, 799–805.

Friedman,N. *et al.* (2000) Using Bayesian networks to analyze expression data. *J. Comput. Biol.*, **7**, 601–620.

Fuller,T.F. *et al.* (2007) Weighted gene coexpression network analysis strategies applied to mouse weight. *Mamm. Genome*, **18**, 463–472.

Gomez,B.P. *et al.* (2007) Human X-Box binding protein-1 confers both estrogen independence and antiestrogen resistance in breast cancer cell lines. *FASEB J.*, **21**, 4013–4027.

Gompel,A. *et al*. (2000) Hormonal regulation of apoptosis in breast cells and tissues. *Steroids*, **65**, 593–598.

Heckerman,D. *et al.* (2000) Dependency networks for inference, collaborative filtering, and data visualization. *J. Mach. Learn. Res.*, **1**, 49–75.

Hood,L. *et al*. (2004) Systems biology and new technologies enable predictive and preventative medicine. *Science*, **306**, 640–643.

Howell,A. (2006) Pure oestrogen antagonists for the treatment of advanced breast cancer. *Endocr. Relat. Cancer*, **13**, 689–706.

Husmeier,D. (2003) Sensitivity and specificity of inferring genetic regulatory interactions from microarray experiments with dynamic Bayesian networks. *Bioinformatics*, **19**, 2271–2282.

Iwakoshi,N.N. *et al.* (2003) The X-box binding protein-1 transcription factor is required for plasma cell differentiation and the unfolded protein response. *Immunol. Rev.*, **194**, 29–38.

Kim,H. *et al.* (2006) Unraveling condition specific gene transcriptional regulatory networks in Saccharomyces cerevisiae. *BMC Bioinformatics*, **7**, 165.

Kitano,H. (2002) Systems biology: a brief overview. *Science*, **295**, 1662–1664.

Kostka,D. and Spang,R. (2004) Finding disease specific alterations in the co-expression of genes. *Bioinformatics*, **20**, i194–i199.

Kuo,M.T. (2007) Roles of multidrug resistance genes in breast cancer chemoresistance. In *Breast Cancer Chemosensitivity*. Springer, Berlin, pp. 23–30.

Li,C. and Li,H. (2008) Network-constrained regularization and variable selection for analysis of genomic data. *Bioinformatics*, **24**, 1175–1182.

Li,H. *et al.* (2008) Inferring regulatory networks. *Front. Biosci.*, **13**, 263–275.

Liao,J.C. *et al*. (2003) Network component analysis: Reconstruction of regulatory signals in biological systems. *Proc. Natl Acad. Sci. USA*, **100**, 15522–15527.

Lin,C.Y. *et al.* (2004) Discovery of estrogen receptor alpha target genes and response elements in breast tumor cells. *Genome Biol.*, **5**, R66.

Liu,C.C. *et al*. (2006) Topology-based cancer classification and related pathway mining using microarray data. *Nucleic Acids Res.*, **34**, 4069–4080.

Liu,Y. *et al.* (2006) Genome wide profiling of human embryonic stem cells (hESCs), their derivatives and embryonal carcinoma cells to develop base profiles of U.S. Federal government approved hESC lines. *BMC Dev. Biol.*, **6**, 20.

Luscombe,N.M. *et al*. (2004) Genomic analysis of regulatory network dynamics reveals large topological changes. *Nature*, **431**, 308–312.

Meinshausen,N. and Buhlmann,P. (2006) High-dimensional graphs and variable selection with the Lasso. *Ann. Stat.*, **34**, 1436–1462.

Pratt,M.A.C. *et al.* (2003) Estrogen withdrawal-induced NF-kappa B activity and Bcl-3 expression in breast cancer cells: roles in growth and hormone independence. *Mol. Cell. Biol.*, **23**, 6887–6900.

Qiu,P. *et al*. (2005) Ensemble dependence model for classification and prediction of cancer and normal gene expression data. *Bioinformatics*, **21**, 3114–3121.

Qiu,P. *et al.* (2007) Dependence network modeling for biomarker identification. *Bioinformatics*, **23**, 198–206.

Rangel,C. *et al*. (2004) Modeling T-cell activation using gene expression profiling and state-space models. *Bioinformatics*, **20**, 1361–1372.

Riggins,R.B. *et al*. (2003) Synergistic promotion of c-Src activation and cell migration by Cas and AND-34/BCAR3. *J. Biol. Chem.*, **278**, 28264–28273.

Riggins,R.B. *et al.* (2005) Antiestrogens, aromatase inhibitors, and apoptosis in breast cancer. In *Vitamins and Hormones - Advances in Research and Applications*. Vol. 71. Elsevier Academic Press Inc, San Diego, pp. 201–237.

Riggins,R.B. *et al.* (2007) Pathways to tamoxifen resistance. *Cancer Lett.*, **256**, 1–24.

Sanna,M.G. *et al.* (2002) IAP suppression of apoptosis involves distinct mechanisms: the TAK1/JNK1 signaling cascade and caspase inhibition. *Mol. Cell. Biol.*, **22**, 1754–1766.

Sato,N. *et al*. (2003) Molecular signature of human embryonic stem cells and its comparison with the mouse. *Dev. Biol.*, **260**, 404–413.

Schrecengost,R.S. *et al.* (2007) Breast cancer antiestrogen resistance-3 expression regulates breast cancer cell migration through promotion of p130(Cas) membrane localization and membrane ruffling. *Cancer Res.*, **67**, 6174–6182.

Segal,E. *et al.* (2003) Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. *Nat. Genet.*, **34**, 166–176.

Shmulevich,I. *et al*. (2002) Probabilistic Boolean networks: a rule-based uncertainty model for gene regulatory networks. *Bioinformatics*, **18**, 261–274.

Somai,S. *et al.* (2003) Antiestrogens are pro-apoptotic in normal human breast epithelial cells. *Int. J. Cancer*, **105**, 607–612.

Sun,Y. *et al.* (2006) Mechanisms controlling embryonic stem cell self-renewal and differentiation. *Crit. Rev. Eukaryot. Gene Expr.*, **16**, 211–231.

Sun,Y. *et al.* (2007) Cross-species transcriptional profiles establish a functional portrait of embryonic stem cells. *Genomics*, **89**, 22–35.

Tibshirani,R. (1996) Regression shrinkage and selection via the Lasso. *J. R. Stat. Soc. Ser. B Methodol.*, **58**, 267–288.

Tozlu,S. *et al.* (2006) Identification of novel genes that co-cluster with estrogen receptor alpha in breast tumor biopsy specimens, using a large-scale real-time reverse transcription-PCR approach. *Endocr. Relat. Cancer*, **13**, 1109–1120.

Van Agthoven,T. *et al.* (2006) Functional identification of genes causing estrogen independence. *Breast Cancer Res. Treat.*, **100**, S37–S37.

Van den Bulcke,T. *et al.* (2006) SynTReN: a generator of synthetic gene expression data for design and analysis of structure learning algorithms. *BMC Bioinformatics*, **7**, 43.

Viatour,P. *et al.* (2003) NF-kappa B2/p100 induces Bcl-2 expression. *Leukemia*, **17**, 1349–1356.

Wang,D.Y. *et al.* (2004) Identification of estrogen-responsive genes by complementary deoxyribonucleic acid microarray and characterization of a novel early estrogen-induced gene: EEIG1. *Mol. Endocrinol.*, **18**, 402–411.

Watson,M. (2006) CoXpress: differential co-expression in gene expression data. *BMC Bioinformatics*, **7**, 509.

Wei,Z. and Li,H. (2007) A Markov random field model for network-based analysis of genomic data. *Bioinformatics*, **23**, 1537–1544.

Zhan,M. (2008) Genomic studies to explore self-renewal and differentiation properties of embryonic stem cells. *Front Biosci.*, **13**, 276–283.

Zhan,M. *et al*. (2005) Conservation and variation of gene regulation in embryonic stem cells assessed by comparative genomics. *Cell Biochem. Biophys.*, **43**, 379–405.

Zhao,P. and Yu,B. (2006) On model selection consistency of Lasso. *J. Mach. Learn. Res.*, **7**, 2541–2563.

Zou,H. and Hastie,T. (2005) Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **67**, 301–320.

Journal of Steroid Biochemistry & Molecular Biology 114 (2009) 8–20

Contents lists available at ScienceDirect

Journal of Steroid Biochemistry and Molecular Biology

journal homepage: www.elsevier.com/locate/jsbmb

Gene network signaling in hormone responsiveness modifies apoptosis and autophagy in breast cancer cells☆

Robert Clarke[a,b,*], Ayesha N. Shajahan[a], Rebecca B. Riggins[a], Younsook Cho[a], Anatasha Crawford[a], Jianhua Xuan[c], Yue Wang[c], Alan Zwart[a], Ruchi Nehra[a], Minetta C. Liu[a]

[a] Department of Oncology and Lombardi Comprehensive Cancer Center, Georgetown University School of Medicine, Washington, DC, USA
[b] Department of Physiology & Biophysics, Georgetown University School of Medicine, Washington, DC, USA
[c] Department of Electrical and Computer Engineering, Virginia Polytechnic Institute and State University, Arlington, VA, USA

ARTICLE INFO

Article history:
Received 25 November 2008
Accepted 31 December 2008

Keywords:
Antiestrogen
Autophagy
Apoptosis
Breast cancer
Cell signaling
Endoplasmic reticulum
Estrogens
Gene networks
Unfolded protein response

ABSTRACT

Resistance to endocrine therapies, whether *de novo* or acquired, remains a major limitation in the ability to cure many tumors that express detectable levels of the estrogen receptor alpha protein (ER). While several resistance phenotypes have been described, endocrine unresponsiveness in the context of therapy-induced tumor growth appears to be the most prevalent. The signaling that regulates endocrine resistant phenotypes is poorly understood but it involves a complex signaling network with a topology that includes redundant and degenerative features. To be relevant to clinical outcomes, the most pertinent features of this network are those that ultimately affect the endocrine-regulated components of the cell fate and cell proliferation machineries. We show that autophagy, as supported by the endocrine regulation of monodansylcadaverine staining, increased LC3 cleavage, and reduced expression of p62/SQSTM1, plays an important role in breast cancer cells responding to endocrine therapy. We further show that the cell fate machinery includes both apoptotic and autophagic functions that are potentially regulated through integrated signaling that flows through key members of the BCL2 gene family and beclin-1 (BECN1). This signaling links cellular functions in mitochondria and endoplasmic reticulum, the latter as a consequence of induction of the unfolded protein response. We have taken a seed-gene approach to begin extracting critical nodes and edges that represent central signaling events in the endocrine regulation of apoptosis and autophagy. Three seed nodes were identified from global gene or protein expression analyses and supported by subsequent functional studies that established their abilities to affect cell fate. The seed nodes of nuclear factor kappa B (NFκB), interferon regulatory factor-1 (IRF1), and X-box binding protein-1 (XBP1) are linked by directional edges that support signal flow through a preliminary network that is grown to include key regulators of their individual function: NEMO/IKKγ, nucleophosmin and ER respectively. Signaling proceeds through BCL2 gene family members and BECN1 ultimately to regulate cell fate.

1. Introduction

Over 40,000 American women die of breast cancer each year [1]; incidence is broadly similar across the European Union when considered as a percentage of the population. In 2008, over 178,000 women will be diagnosed with invasive breast cancer in the U.S., almost 70% of which will be estrogen receptor-α positive (ER+; HUGO Gene Symbol = ESR1) [2,3]. The percentage of ER+ sporadic breast cancers increases linearly with age but even in premenopausal cases the proportion is high; 62% at age ≤35 and 72% by age 49 [2–4]. Data from randomized trials and meta-analyses clearly show that all breast cancer patients derive a statistically significant survival benefit from adjuvant chemotherapy, and that all hormone receptor positive breast cancer patients benefit from adjuvant endocrine therapy [5–9]. For postmenopausal women, the benefit from adjuvant Tamoxifen (TAM) is comparable to that seen for cytotoxic chemotherapy. While 5 years of adjuvant TAM produces a 26% proportional reduction in mortality [8], many ER+ tumors eventually recur [10]. Since advanced ER+ breast cancer largely remains an incurable disease, resistance to endocrine therapies is a significant clinical problem.

Endocrine therapy is administered as an antiestrogen (AE) like Tamoxifen (TAM) or Fulvestrant (FAS; Faslodex; ICI 182,780), or as an aromatase inhibitor (AI) such as Letrozole or Exemestane. It is less toxic and potentially more effective therapy in the management of hormone-dependent breast cancers. Antiestrogens, and TAM in

☆ Lecture presented at the '18th International Symposium of the Journal of Steroid Biochemistry and Molecular Biology', 18–21 September 2008, Seefeld, Tyrol, Austria.

* Corresponding author at: Room W405A Research Building, Department of Oncology, Georgetown University School of Medicine, 3970 Reservoir Rd, NW, Washington, DC 20057, USA. Tel.: +1 202 687 3755; fax: +1 202 687 7505.
E-mail address: clarker@georgetown.edu (R. Clarke).

0960-0760/$ – see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jsbmb.2008.12.023

particular, have been the "gold standard" first line endocrine therapy for over 30 years [11], clinical experience with this drug likely exceeding over 15 million patient years [10]. TAM increases both disease free and overall survival from early stage breast cancer, and it also reduces the incidence of invasive and noninvasive breast cancer in high-risk women [8,9]. Raloxifene, another antiestrogen, is effective in reducing the rate of postmenopausal bone loss from osteoporosis as well as the rate of invasive breast cancer [12]. Newer antiestrogens such as FAS show significant activity relative to TAM and some AIs [13,14]. Third generation AIs are now widely accepted as viable alternatives to AEs for first line endocrine therapy in postmenopausal women with metastatic disease; overall response rates are generally greater for AIs [15]. Importantly, Tamoxifen is the only single agent with demonstrated efficacy in both premenopausal and postmenopausal women with invasive breast cancer. Other AEs and all of the AIs require the complete cessation of ovarian function.

Of current interest is identification of the optimum choice and scheduling of AEs and AIs. Evidence clearly shows improvements in disease free survival for combined adjuvant therapy (an AI and an AE usually given sequentially) over single agent TAM [16–20]. However, the ability of AIs to induce a significant improvement in overall survival compared with 5 years of TAM alone is uncertain [15]. In terms of metastatic disease, recent data imply that response rates with an AI are either equivalent with or higher than with TAM [21,22]. Given the increasing number of endocrine treatment options, there is a clear need to optimize the selection and scheduling of agents for both early stage and advanced disease. Whichever way these controversies are eventually resolved, it is clear that both AIs and AEs will remain as key modalities in the management of ER+ breast cancers. Unfortunately, the inability of endocrine therapies to cure many women with ER+ disease will also remain.

1.1. Endocrine resistance: receptor phenotypes

Several resistance phenotypes are evident from both experimental models and clinical observations. The two primary receptor phenotypes are ER+ and ER−. These receptor-based phenotypes have been further stratified by addition of the estrogen-regulated receptor for progesterone (PGR; HUGO Gene Symbol = PGR). The degree of treatment benefit from endocrine therapy varies according to receptor phenotype. For example, approximately 75% of ER+/PGR+, 33% of ER+/PGR−, and 45% of ER−/PGR+ cases of metastatic breast cancer respond to TAM [10]. Endocrine responses in truly ER− tumors are probably relatively rare and of uncertain relevance, as they most likely reflect incorrect assessments of what may be very low ER and/or PGR expression values. Data from the Early Breast Cancer Trialists' Collaborative Group meta-analyses show that TAM therapy generates a non-significant 6% reduction in the 10-year risk of recurrence. A non-significant increase in the risk of death from any cause in patients with ER− breast cancer also was reported [8,9]. The real value of PGR, which is the only modification to this clinical prediction scheme for directing endocrine therapy to become routine in over 30 years (the value of directing endocrine therapy based on HER2 is still controversial), is largely limited to ER− tumors. It is general practice in the United States to treat all ER+ and/or PR+ invasive breast tumors with endocrine therapy. However, it remains impossible to predict whether an individual patient will receive benefit from treatment and the magnitude or duration of any benefit. Better predictors of each individual patient's endocrine responsiveness are clearly needed.

1.2. Endocrine resistance: pharmacological phenotypes

Several pharmacological phenotypes have been identified in experimental models of either human breast cancer cells growing *in vitro* or of xenografts in immune-deficient rodents [10]. These phenotypes include (i) estrogen-independent (which appears equivalent to AI resistance but is not so for antiestrogen resistance [23]—some breast cancers can become resistant to an AE but still respond to an AI and *vice versa*); (ii) estrogen-inhibited (recently identified in MCF-7 models [24]); (iii) TAM-stimulated (identified first in MCF-7 xenografts [25,26]); TAM-unresponsive but FAS sensitive [27] (identified first in MCF-7 models and subsequently observed in clinical trials [13]); TAM and FAS crossresistant [28] (perhaps this is truly antiestrogen crossresistant and it is seen both clinically in patients and experimentally in MCF-7 models [13,29]). Other variations on these phenotypes likely occur but are beyond the scope of our discussion.

1.3. Clinical evidence for the prevalence of pharmacological resistance phenotypes

Obtaining direct clinical evidence for the prevalence of each of the pharmacological resistance phenotypes is challenging. We have previously noted the utility of applying clinical responses to TAM withdrawal in metastatic breast cancer as one means to define, at least in broad terms, the likely relevance of a series of pharmacological phenotypes [29]. This approach is somewhat limited, as the number of cases across all studies is modest ($n = 241$). Furthermore, TAM withdrawal responses cannot readily distinguish between TAM-stimulation and estrogen-inhibition because each should predict for a clinical benefit. The latter would induce a benefit because many breast cancers contain significant concentrations of 17β-estradiol, independent of both menopausal and ER/PGR status [10], sufficient to produce the estrogen-inhibited phenotype [24]. Indeed, the superiority of AIs over TAM in inducing clinical response strongly implies that over 75% of ER+/PGR+, at least 50% of all ER+ breast cancers irrespective of PGR expression, and 45% or more of ER−/PGR+ breast tumors are probably driven by adequate access to estrogen.

In our prior assessment, almost 9% of patients received an overall clinical response to TAM withdrawal (partial responses + complete responses). When disease stabilizations were included we estimated that less than 20% of patients received clinical benefit [29], suggesting that the sum of TAM-stimulated plus estrogen-inhibited clinical phenotypes may not account for the majority of resistant phenotypes in women. Of course, given the number of ER+ breast cancers arising every year, these phenotypes are relevant to a notable number of women. The major response to TAM withdrawal was clinically detectable disease progression – greater than 80% of cases – strongly implicating unresponsiveness as the primary clinical resistance mechanism to TAM. Whether these breast cancers are fully crossresistant to all endocrine therapies, or retain sensitivity to AIs, cannot be determined from this simple analysis.

Nomura et al. [30] took a different approach and assessed the responsiveness to estrogen and TAM in short-term primary cell cultures of $n = 153$ ER+ breast cancer biopsies. This approach allowed the authors to separate the various pharmacological phenotypes; approximately 7% of ER+ primary cultures were stimulated by TAM and almost 3% were inhibited by physiological concentrations of estradiol—notably close to our estimate of 9% for the sum of these two clinical phenotypes.

It is important here to separate responses to physiological estrogens from those produced by pharmacological estrogen therapy. High dose estrogen therapy was used prior to the advent of TAM. As with all endocrine therapies, approximately 30% of all breast cancers (receptor status was not available when most of these studies were done) responded [31,32]. Side effects were unfavorable, probably explaining the switch to TAM that also induces responses in approximately 30% of all breast cancers (when receptor status is not considered). It is also likely that the mechanisms of action of pharmacological and physiological dose estrogens differ. Over 15 years

ago, we were the first to show that pharmacological concentrations of both estradiol and TAM induce changes in the membrane fluidity of breast cancer cells and that this correlates with changes in cell growth [33]. It is unlikely that membrane fluidity changes are major contributors to the action, either prosurvival or prodeath, of physiological estrogen exposures but they likely do contribute to the prodeath effects of pharmacological exposures.

2. Cell fate in the context of endocrine responsiveness

Therapeutic strategies for breast cancer generally aim to alter the balance between cell death and cell survival such that cancer cells (but ideally not normal cells) die. However, endocrine therapies consistently also induce a notable growth arrest in sensitive tumors. The relative importance of growth arrest and cell death remains unclear. To explore this issue, we will first discuss the forms of cell death and then compare the potential for cell death and cell growth arrest to contribute to endocrine responsiveness.

Cell death pathways include signaling to apoptosis, autophagy, mitotic catastrophe, necrosis, and senescence. Late events in cell death are reasonably well defined at the molecular (such as PARP cleavage) and cellular levels (including DNA disintegration). However, knowledge of the regulatory signaling upstream of these events, and how this signaling is integrated and processed, is now known to be incomplete. Mitochondrial function and integrity, regulated in part by BCL2 family members, are central to several forms of cell death [34–36].

2.1. Apoptosis

Apoptosis is a programmed cell death defined by morphological criteria related to organized chromatin condensation and fragmentation of the cell nucleus, accompanied by cleavage of DNA, formation of apoptotic bodies, cell shrinkage, and ruffling of the cell membrane [35,37,38]. Two major pathways are involved. The *intrinsic (mitochondrial) pathway* is regulated by the proapoptotic and antiapoptotic BCL2 family members; this pathway involves changes in mitochondrial membrane permeability (MMP), release of cytochrome c, exposure of phosphatidylserine on the outer leaflet of the plasma membrane, and the eventual loss of plasma membrane integrity [39]. The *extrinsic (cell surface receptor) pathway* is dependent upon extracellular signals including tissue necrosis factor-α (TNFα), Fas ligand, and TNF-related ligand TRAIL [37,38]. The intrinsic and extrinsic pathways activate caspases, the "executioners" of apoptosis, which cleave DNA and catabolize the cytoskeleton. Apoptosis is not a discrete process and occurs over time—early (4–18 h), middle (18–36 h), and late stages (≥36 h) are often described based largely on data from cell culture models. Changes in specific BCL2 family members (early events that can precede changes in MMP), changes in MMP, and the exposure of phosphatidylserine are generally interpreted as representing early-to-middle apoptosis. Cytoplasmic cytochrome c release from mitochondria, changes in propidium iodide staining, increased terminal transferase dUTP nick end labeling (TUNEL) and cleavage of the DNA repair enzyme PARP-1 are associated with late apoptosis or necrosis [35].

2.2. Autophagy

Autophagy is a lysosomal pathway where cytoplasmic contents are degraded by double/multi-membrane vacuoles or autophagosomes, normally resulting in the removal of defective or damaged organelles, *e.g.*, mitochondria. A better understanding of the regulation of autophagy has recently begun to emerge; key regulators are now known to include BCL2 family members [40,41] and their interacting proteins such as beclin-1/ATG6 (BECN1) [42]. BCL2 antiapoptotic proteins can block autophagy by inhibiting BECN1 [36]. Since monoallelic loss of the BECN1 locus is seen in >40% of breast cancers [43] (and in MCF-7 cells), modulating BCL2 may be an effective mechanism for regulating BECN1-activated autophagy. Autophagy can be identified by the absence of marginated nuclear chromatin, the presence of cytoplasmic vacuoles using transmission electron microscopy or monodansylcadaverine [44,45], cleavage of the LC3B protein [46,47], and regulation of the p62/SQSTM1 protein [48]. Early events in autophagy may be reversible; later events may (or appear to) share mechanisms with other cell death pathways. For example, cleavage of ATG5 by caplain [49] or upregulation of BID [41] can cause a switch from autophagy to apoptosis.

Paradoxically, autophagy can act as a cell survival mechanism when extracellular nutrients or growth factors are limited, or as an alternative cell death pathway to apoptosis [50]. Prosurvival outcomes likely reflect an adequate adjustment to stress, with energy/nutrients recovered from the organelles "digested" in the autophagosomes. Prodeath outcomes may arise when the self-digestion of autophagy leads to such a loss of organelles that the cell can no longer survive. In cancer cells, autophagy induction can accelerate cell death [51–55] or promote cell survival [56–58], independently or in response to treatment with cytotoxic agents.

2.3. Mitotic catastrophe

Faulty DNA structure checkpoints, or the spindle assembly checkpoint, are key components of this form of cell death [59,60]. Disruption of the normal segregation of many chromosomes results in rapid cell death [59]. When this cell death does not occur, the cell can divide asymmetrically and produce aneuploid daughter cells [61] that can become neoplastic [59,61]. Thus, mitotic catastrophe is characterized by multinucleation.

2.4. Necrosis

Necrosis is a chaotic process marked by cellular edema, vacuolization of the cytoplasm, breakdown of the plasma membrane, and an associated inflammatory response caused by the release of cell contents into the surroundings. Increased permeability to trypan blue or other vital dyes, in the absence of organized chromatin condensation and DNA fragmentation, is characteristic of necrosis [44,62].

2.5. Senescence

Senescent cells are characteristically enlarged, flattened with vacuoles and a large nucleus, be come permanently cell cycle arrested and unresponsive to mitogenic stimuli and express β-galactosidase [45,63]. Normally, as telomerase activity falls over time, successive telomere shortening limits proliferation and leads to "cellular senescence" or "mortality stage 1 (M1)". Inactivation of p53 can by bypass M1 growth arrest, producing critically short telomeres and massive cell death called "mortality stage 2 (M2)" or "crisis" [64].

2.6. Endocrine-induced cell death in breast cancer

Precisely how breast cancer cells die following estrogen withdrawal (or AI treatment) or AE treatment is unclear. Senescence may not be the dominant mechanism, since this process frequently involves DNA damage and p53 activation [38,45] but breast cancer cells respond to AEs and to estrogen withdrawal even if they have mutated p53 [35,65]. While apoptosis is clearly implicated [65–68], some of the apoptosis endpoints in prior studies may not

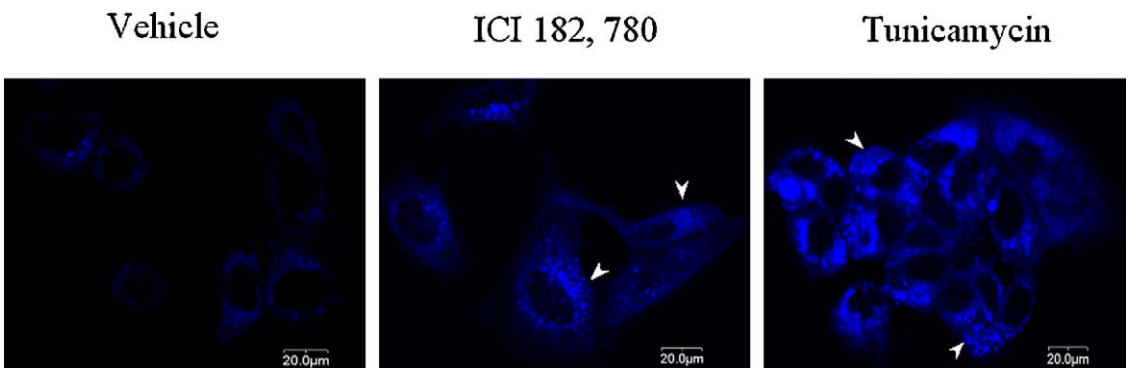


Fig. 1. Autophagy is enhanced upon FAS treatment in ER+ breast cancer cell lines. MCF-7 cells were treated with FAS (ICI 182,780), the endoplasmic reticulum stress and autophagy inducer tunicamycin (TUN), or ethanol control (vehicle) prior to staining with monodansylcadaverine (MDC). Increased MDC staining indicates that autophagy has been induced.

distinguish among earlier events more closely implicated with signaling initiated through autophagy. Autophagy has been implicated in response to endocrine therapy [69–71] and we also see the induction of significant autophagy associated with endocrine therapies.

Fig. 1 shows our ability to detect significant changes in the number of autophagosomes as measured by an increase in the presence of cytoplasmic vacuoles identified by monodansylcadaverine staining [44,45] (Fig. 1), increased cleavage of the LC3 protein [46,47], and reduced expression of p62/SQSTM1 [48,72–74] (Fig. 2). We have previously shown, as have others, that AE treatment and estrogen withdrawal are also accompanied by increases in the level of apoptosis and growth arrest in sensitive cells. Indeed, when restoring AE sensitivity in resistant cells we frequently see that sensitivity is reflected in the restoration of an ability of the antiestrogen (or estrogen withdrawal) to both increase apoptosis and reduce proliferation [75,76]. As shown in Figs. 1 and 2, and consistent with other reports [69–71], prodeath autophagy also is associated with the growth inhibitory effects of endocrine therapies in breast cancer cells. Thus in experimental models, cells responding to endocrine therapies concurrently experience an increase in cell growth arrest accompanied by both apoptosis and a prodeath autophagy.

2.7. Proliferation, cell death, and endocrine responsiveness

One of the most consistent observations in both experimental models *in vitro* and *in vivo* and in clinical specimens is the ability of endocrine therapies to induce a profound growth arrest in sensitive breast cancer cells. However, the relative importance of increased cell death compared with reduced proliferation is not entirely clear. In most endocrine sensitive experimental models, growth arrest and cell death concurrently occur and both clearly

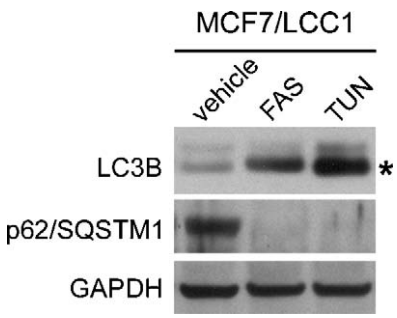


Fig. 2. Autophagy is enhanced upon FAS treatment in ER+ breast cancer cell lines. MCF7/LCC1 cells were treated with FAS, TUN, or vehicle prior to lysis and immunoblotting using standard procedures. Increased LC3BII (asterisk) and decreased p62/SQSTM1 expression both indicate that autophagy has been induced.

contribute to the ability of endocrine therapies to affect changes in anchorage-dependent cell number, anchorage-independent colony formation, or tumorigenesis over time [27,77,78]. Less clear is their relative contribution in driving clinical responses to endocrine therapies. Growth arrest appears to be readily detected in breast tumors responding to endocrine therapy. Less clear is the ability to detect robust changes in apoptosis. Some investigators do [79], and some do not [66], see an association of apoptosis or a molecular maker(s) of apoptosis with clinical response. The latter is in marked contrast to studies in experimental models. For some studies, response is related to molecular markers of apoptosis such as BCL2 [79] or the FasL:Fas ratio [80]. Notably, expression of the anti-apoptotic molecule BCL2 is reduced in responsive breast tumors by 3 months of TAM treatment [79], while in breast tumors that remain after TAM therapy BCL2 expression is elevated [81]. However, as noted above, BCL2 can affect both an apoptotic and autophagic cell death and its measurement alone is likely a poor predictor of any specific cell death mechanism.

If cell death does not occur in clinical breast cancer this observation clearly requires explanation. Several possible explanations exist—in the absence of compelling experimental/clinical data supporting or eliminating these explanations we make no assessment at this time on their relative merits. Firstly, it should be noted that measures of apoptosis are usually the primary endpoints for assessing rates of cell death. Our previously published results, the data in Figs. 1 and 2, and the work of others [69–71] show that estrogen withdrawal or antiestrogens increase both the rates of apoptosis and autophagy in breast cancer models responding to treatment. We interpret this as a prodeath autophagy in sensitive cells, consistent with other reports [69–71]. It remains unclear whether autophagy or apoptosis dominates as the cell death mechanism or whether this varies among different breast cancer cells. Measuring apoptosis may be the wrong measure of cell death in tumors, or it may be an inadequate measure if it represents only some proportion of cells that die through this process. Secondly, apoptosis is often considered to comprise early, mid and late stages, and an irreversible commitment to cell death may not be robustly associated with endpoints other than those definitively reflecting late stage apoptosis. A measure of apoptosis that is not robustly associated with ultimate cell death could provide an incomplete assessment of the rate or extent of cell death. Thirdly, if the timing of apoptosis is as fast in patient tumors as it is *in vitro*, measurements taken before 24–36 h and/or after 36–48 h could miss many of the key events. The most sensitive cells would have been through apoptosis and be already dead and gone, and the rate of apoptosis could have returned to the basal level. Fourthly, duration of the apoptotic response may differ between basal apoptosis and drug-induced apoptosis. If drug-induced apoptosis leads to a more rapid death, the number of cells processing though apoptosis could increase

without any detectable change across time in the apparent rate of apoptosis.

Finally, a reduction in cell proliferation alone could be sufficient to account for some shrinkage of tumor size, as the rate of cell replacement might no longer be sufficient to account for cell loss from either a basal rate of cell death and/or loss to migration and metastasis. However, unless almost all growth arrested cells also undergo some form of cell death, it is unclear why growth arrest alone should lead to large and relatively rapid reductions in tumor size (over several weeks compared with often many years of presumably much longer growth prior to clinical detection and treatment). Growth arrest alone may be sufficient to account for good responses in some tumors, particularly where there is a high basal rate of cell death. However, it is not immediately clear how this applies to tumors with an inherently low rate of proliferation, whether because the growth fraction is large but cycling slowly or the growth fraction is small but proliferating rapidly. This is an area where mathematical modeling could be particularly useful, since it could compare the effect sizes needed for relative changes in proliferation and cell death to affect predicted overall tumor size over time.

While there is currently no definitive understanding of the primary cell death mechanisms in either experimental models or in breast tumors in women, or of the relative importance of endocrine therapy-induced changes in proliferation compared with cell death, there are potentially important implications for the underlying biology of the cancer cells. If the primary driver of response as seen in tumor shrinkage is a reduction in proliferation, this will leave many cells alive and still metabolically active. Surviving cells have the ability to adapt to the endocrine-induced stress and eventually overcome the proliferative blockade and grow—they will become resistant. This process seems unlikely to occur in many of those women who receive the clear long term benefit of a significant reduction in the risk of death [8,9].

Whether it is the growth arrested but surviving cells that eventually become resistant is unknown but it is certainly an intuitively satisfying hypothesis. Moreover, this hypothesis is supported by the ability to take sensitive cells in culture, expose them for prolonged periods to either estrogen withdrawal or AE treatment, and eventually induce an acquired resistant phenotype [27,28,77,82]. This process is accompanied by a profound and prolonged period of growth arrest prior to the emergence of resistant cells, a pattern consistent with the clinical progress of the disease in tumors that initially respond to therapy but that eventually recur—often a decade or more after the initiation of TAM treatment.

3. Molecular signaling and resistance

The precise mechanisms of resistance to an AE and/or an AI remain unclear, reflecting an incomplete understanding of the signaling affecting cell proliferation, survival, and death and their hormonal regulation in breast cancer. We have previously reviewed the mechanisms of resistance to AEs and to estrogen deprivation elsewhere in some detail [10,23,29], so we focus here on the molecular signaling aspects of resistance and how these may be integrated and explored using emerging technologies. We will focus primarily on signaling to cell death—signaling to regulate proliferation in the context of endocrine responsiveness will be the subject of a separate review.

The primary technologies that have matured sufficiently to enable global approaches to network modeling include gene expression microarrays, ChIP-on-chip, SNP chips, high-throughput DNA sequencing, and array CGH. Each of these technologies has reached a high level of maturity, and each is characterized by the generation of very high dimensional data on each sample whether the read-out be genomic or transcriptomic data; this also is true of the emerging high-throughput proteomic technologies. The remarkable volume of data, and the diversity of biological information that informs the interpretation of these data, has begun to transform the fields of biostatistics, computer science, and bioinformatics. However, the properties of these datasets are often not fully understood nor are the challenges these properties provide for data analysis and network modeling. Readers interested in exploring some of these challenges can read recent reviews [83,84]. Here we will address briefly several approaches to the use of these data for network modeling.

3.1. A network signaling hypothesis of endocrine responsiveness

Estrogen-independence and AE resistance are complex phenotypes and both genomic and non-genomic activities are implicated [10,33,85]. We consider it unlikely that endocrine resistance in ER+ tumors is driven by a single gene/signaling pathway. Unlike many previous single gene/pathway studies, our central hypothesis invokes a gene network that confers diversity and redundancy in signaling [10,86]. The cell death/survival network incorporates specific signaling as affected by estrogen and AE modification of ERα function. Thus, AEs regulate this network differently than other agents such as cytotoxic drugs.

Signaling leads first to the reversible initiation of several cell death/survival signaling pathways within the network. The irreversible machinery of cell destruction is activated at some later point. This machinery may induce common outcomes – such as activation of effector caspases and DNA/plasma membrane disintegration – independent of the early specific initiating signals. Hence, we envision multiple concurrent signals processing through this network, some prosurvival and some prodeath, with cell fate reflecting the dominant signaling. In endocrine resistant cells, endocrine regulation and/or function of components of this network are changed and prodeath signals are either no longer induced or dominant.

This cell fate signaling network hypothesis is intuitively logical and certainly testable. Evidence that cells induce prosurvival signaling in an attempt to circumvent stressors implies that some cells are successful and ultimately survive whereas others are unsuccessful and die. Thus, the balance between prosurvival and prodeath signaling is likely the final arbiter of cells fate [83]. While this remains an area of active investigation, we first discuss the basic principles of network modeling and then provide an example of a seed-gene network of endocrine-regulated signaling in endocrine responsiveness.

3.2. Basic concepts of gene networks

Cellular signaling occurs more in the context of interactive networks than through linear pathways [83]. The basic topology of a network is defined by nodes (genes/proteins) and their interconnections (edges). Interconnections are multi-faceted and include one-to-one, one-to-many, or many-to-one relationships, and feed-forward or feed-back loops. The dynamic activity of a network is constrained by the various forms of interactions, and the network behaves only in certain ways and controlled manners in response to changing cellular conditions or external stimuli [87]. While often built solely from gene expression microarray data, these data are high dimensional and contain spurious correlations that can confound simple solutions for network building [83,84]. Relevant events also occur in the genome and proteome, some of which can affect the transcriptome. For example, a transcription factor (TF) may be activated by phosphorylation and bind to responsive elements in the genome but the regulation of its downstream targets is seen in the transcriptome [83]. An example of this relationship is

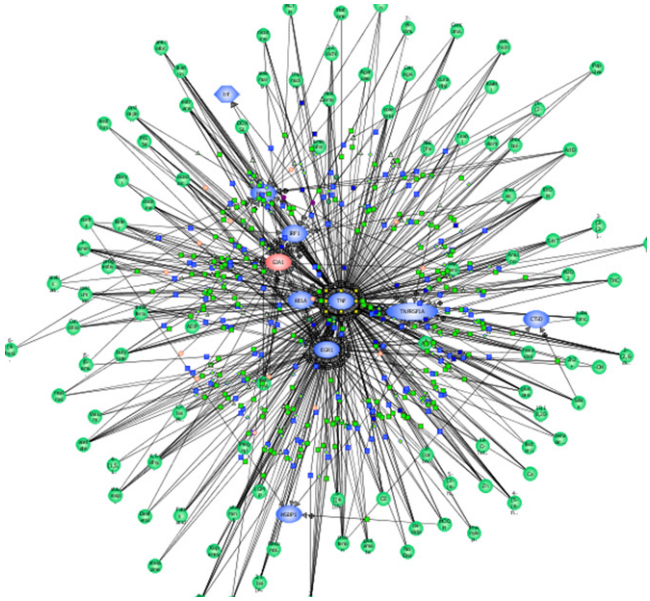


Fig. 3. Illustration of the complex and challenging nature of pathway analysis. Genes identified as being differentially expressed in resistant MCF7/LCC9 cells by SAGE and gene expression microarray were analyzed by Pathway Architect (Stratagene) to identify relationships *in silico*.

the ligand-independent activation of ERα following its phosphorylation on SER118 by MAPK [88].

Simplistically, there are two basic approaches to network modeling of high dimensional data: top-down and bottom-up. The former is probably the most widely used approach as several accessible commercial software packages are available that make this an easy task to perform without the need for training in biostatistics or bioinformatics. These packages often apply various implementations of gene ontologic and semantic search algorithms that identify cellular functions and pathways to which individual nodes are assigned; these data are then graphically represented.

The solutions produced by several popular top-down algorithms are often characterized by representations of tens-to-hundreds of nodes linked by hundreds-to-thousands of edges, making interpretation challenging (Fig. 3). Whether the algorithms address the confounding properties of high dimensional spaces, such as the curse of dimensionality or the confound of multimodality, or incorporate the critical aspects of cellular context and alleviate the trap of self-fulfilling prophesy, is not clear [83]. Among the additional challenges are the incompleteness of relevant biological knowledge and the annotation error rate in the source databases searched by these algorithms [83]. Nonetheless, these approaches can be useful when carefully applied and their limitations fully understood, and when experts from both the biological and mathematics domains combine expertise to assess the validity of the solutions. Currently, such approaches probably have most to offer in the area of hypothesis generation, rather than in the construction of truly biologically meaningful signal transduction networks.

3.3. The "seed-gene" approach to network modeling

The bottom-up approach is generally referred to as the "seed-gene approach" to network modeling [89]. This approach requires the extraction of a small number of seed genes from within the primary data; these genes are then used to grow the network in several ways. We will not address all the various approaches in this review but provide a few brief examples. Various modeling methods can be applied to find and link adjacent nodes, growing the network *de novo*. Local subnetworks can be identified and overlaid or linked to the initial seed genes. A simple approach is the incorporation of a canonical pathway (which may be a subnetwork in what would be a final and much broader network) when it is known to be relevant in the cellular context under study and where incorporating the nodes and edges of the canonical pathway members is consistent with statistical properties of the growing model topology.

Knowledge of how a gene (node) affects the expression/function of another node provides directional connectivity information that can be applied to the interacting nodes. Transcription networks can be grown (or transcriptional edges between nodes in a network that incorporates other biological knowledge) by linking TFs to their downstream targets. These targets can be predicted using specific algorithms [90–93]; where possible it is preferable to incorporate functional data such as that obtained from ChIP-on-chip arrays [91]. Thus, interacting nodes can be identified along with the directionality of their edges as the seed-gene network is grown.

The most labor intensive approach is to derive experimentally nodes and edges, growing the network using definitive laboratory-derived knowledge. Where additional high-throughput data are already available, such as ChIP-on-chip, this is preferable. Currently, functional data is probably more often obtained one gene at a time, using standard molecular methods such as gene knock-down and overexpression. This laborious approach is becoming supplanted with the emerging functional genomic methods such as siRNA, ribozyme, or antisense libraries that can test experimentally the contribution of hundreds to thousands of genes. These methods enable investigators to extract concurrently nodes that experimentally generate biologically appropriate changes in the phenotype under investigation.

Once seeds and their edges are identified, and functional biological metadata obtained, interactive models can be grown using neural network and other machine learning tools. Several models have been proposed to reveal the behaviors of regulatory networks from gene expression data [22,23] including Boolean networks [24–26], Bayesian networks [27–30], linear additive regulation models [31,32], state-space models (SSMs) [33,34], and recurrent neural networks (RNN) [35,36]. However, these methods use only mRNA expression data to infer networks.

Integrated approaches have been recently proposed to learn transcriptional regulation from various data sources [27,30,37–43]. An iterative search on mRNA expression and ChIP-on-chip data [37], or the incorporation of expression profiles, ChIP-on-chip, and motif data [41] have each been used in yeast to discover transcriptional networks. Several linear models or matrix decomposition methods have also been proposed [43–46]. Network component analysis (NCA) is a notably powerful approach [45] but NCA and these other methods cannot easily infer regulatory networks in biological systems more complex than yeast.

Other limitations exist in network modeling. Complete biological knowledge for topology estimation (node–node edges and directionality), such as high-throughput ChIP-on-chip data or functional data from laboratory experiments, are often not (or only partially) available for human cells. When heterogeneous data sources are integrated for computational inference, the consistency of different data sources is often inadequate or unknown. Topological knowledge also comes from biological experiments, which often contains false positives/negatives that can lead to incorrect network inference.

4. Seed-gene model for cell signaling and the regulation of cell fate

While we continue to develop new methods for network modeling, we have yet to report our modeling approaches to our own expanding data sets. Hence, we will here describe our initial studies

on the use of seed genes and experimental data to construct a simple wiring-diagram of our initial seed-gene network. The inability to induce signaling to irreversible cell death is a central component of drug resistance [94]. Thus, we propose that cells possess a common cell death/survival regulatory decision network of integrated and/or interacting pathways (see above).

Prior to building network models, it is necessary to extract initial nodes (seed genes) from which a network can be built [89]. Since ER is a TF and regulates other functionally relevant TFs that influence endocrine responsiveness and cell fate, selecting a small number of TFs as seed genes is reasonable for network modeling. The full list of *relevant* ER-regulated TFs that may affect cell fate is unknown. Nonetheless, our published data support the central hypothesis that that IRF1 [65,95–97], XBP1 [76,95] and NFκB (RELA) [75,95] are key regulatory nodes or control key modules in this network. Moreover, our experimental data in endocrine sensitive and resistant breast human cancer cells now allow us to map their edges and directionality, in an appropriate cellular context, with some confidence.

4.1. X-box binding protein-1 (XBP1) and the unfolded protein response (UPR)

UPR is a central component of the endoplasmic stress response [98], an adaptive signaling pathway that allows cells to survive the accumulation of unfolded proteins in the endoplasmic reticulum lumen [99]. Initially a compensatory mechanism allowing cells to recover normal endoplasmic reticulum function, a prolonged UPR may induce cell death. UPR, which can be induced by cellular stressors such as hypoxia, is activated by each of three molecular sensors: IRE1α, ATF6, PERK [100]. XBP1's *unconventional* splicing (occurs in the cytosol) by IRE1α is an obligate component in both IRE1α- and ATF6-induced UPR [100,101]. The UPR (initiated by XBP1 splicing by IRE1α) can activate autophagy [102]. Whether this is a prosurvival or prodeath form of autophagy is unknown, since UPR activation also can induce both prodeath and prosurvival outcomes [103].

XBP1 is a transcription factor that belongs to the basic region/leucine zipper (bZIP) family [104,105]. The unspliced form, XBP1(U), has a molecular weight of ~33 kDa and acts as a dominant negative of spliced XBP1 [106,107]. The spliced form, XBP1(S), has a molecular weight of ~54 kDa; splicing removes a 26 bp intron and creates a translational frame-shift. Regulation of transcription by XBP1(S) is a consequence of its homodimers activating specific cAMP response elements (CREs) with a conserved ACGT core sequence GATG**ACGT**G(T/G) NNN(A/T)T—sometimes called the UPR element [103,104,108]. XBP1(S), which is implicated in affecting plasma cell differentiation [109], is essential for fetal survival, neurological development, bone growth, immune system activation, and liver development [110,111]. XBP1 is also rapidly induced in response to estrogen-stimulation [112,113]. Consistent with the work of others [108], we have shown that XBP1(S) can bind to and activate ERα in a ligand-independent manner (Fig. 4).

We have recently shown that XBP1(S) confers E2-independence (effectively an AI resistant phenotype) and AE crossresistance (TAM and FAS crossresistance) in both MCF-7 and T47D human breast cancer cells [76]. This activity appears to be driven primarily by XBP1(S), as introduction of the full-length XBP1 cDNA in either MCF-7 or T47D cells generates predominately the XBP1(S) protein. This observation suggests that the basal activity of IRE1α is already adequate and that XBP1(S) is the rate limiting protein. XBP1 is the only known substrate for the IRE1α endonuclease and only IRE1α can splice mammalian XBP1. Since XBP1 splicing is thought to function primarily within the UPR, breast cancer cells may be primed to respond to multiple stressors by activating a prosurvival induction of UPR.

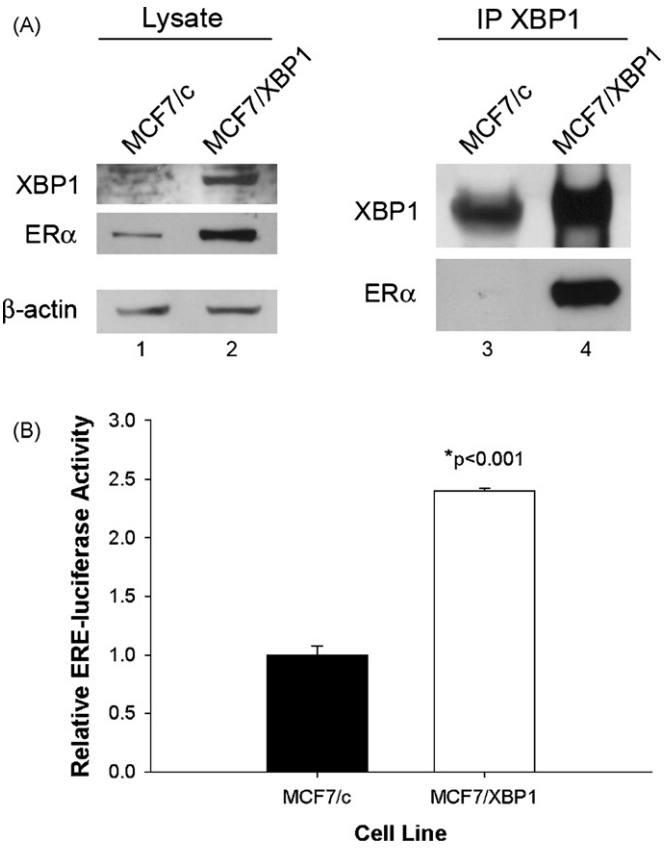


Fig. 4. Physical association of XBP1 and ERα is accompanied by robust ERE-driven transcriptional activity in MCF7/XBP1 cells. (A) MCF-7 cells stably expressing XBP1 cDNA or the empty vector control (c) were treated with FAS or ethanol control (ctrl.) vehicle prior to lysis and immunoblotting (lanes 1 and 2) or co-immunoprecipitation of XBP1 and ERα (lanes 3 and 4) using standard procedures. (B) MCF7/c and MCF7/XBP1 cells were transiently co-transfected with plasmids encoding 3xERE-luciferase and phRLSV40-Renilla for 24 h prior to lysis and promoter–reporter luciferase assay by standard methods. Data are presented as mean relative ERE-luciferase activity ±SE for a representative experiment performed in triplicate, $^{*}p < 0.001$.

4.2. Interferon regulatory factor-1 (IRF1)

RFLP linkage analysis assigned the IRF1 gene to 5q23-31; more definitive studies identified the locus as 5q31.1 [114]. IRF1 was initially identified because of its transcriptional activation of type I interferon (IFN) genes. We first showed the ability of interferons to sensitize breast cancer cells to TAM over 20 years ago [115]. More recently, IRF1 was implicated in T-cell development [116], and it is now known also to coordinate expression of the immunoproteasome [117], to regulate human telomerase activity [118,119], and to regulate key aspects of DNA damage repair [120,121]. Loss of IRF1 increases tumorigenicity in mouse models driven by ras or loss of p53 [122]. These activities may reflect IRF1's ability to signal to apoptosis [123], which can occur in a p53-dependent or -independent manner [120,124], with or without induction of $p21^{cip1}$ [124] or $p27^{kip1}$ [125], and through caspase-1 [120], caspase-3 [96], caspase-7 [96,126], caspase-8 [96,127], and/or FasL [128].

Following our initial observations of IRF1's likely role in breast cancer [129–131] and antiestrogen resistance [129], we confirmed its functional involvement using a dominant negative approach (dnIRF1) [65]. IRF1 and dnIRF1 induce opposing effects on proliferation *in vitro* and tumorigenesis *in vivo* through regulation of caspases-3/7 and caspase-8 activities [96]. These observations are consistent with the effects of inoculating an adenoviral vec-

tor containing IRF1 directly into mouse mammary tumors [132]. While p53-dependent apoptosis occurs in the breast [133], T47D cells express mutant p53 and our data show that intact p53 is not required for the proapoptotic actions of IRF1 [65,96]. In AE sensitive breast cancer cells, inhibition of AE-induced IRF1 activity by dnIRF1 is accompanied by reduced proapoptotic activity [65]. These observations on IRF1 and AE responsiveness have been confirmed and extended by others in both normal [134] and other neoplastic breast cell culture models [135,136]. IRF1, which can signal through both p53-dependent and -independent mechanisms [120,124], provides a new and potentially important signaling molecule for integrating and regulating breast cancer cell survival in response to AEs.

4.3. Nuclear factor kappa B (NFκB)

The NFκB p50/p65 heterodimer complex comprises two homologous proteins; the p50 product of its p105 precursor (NFκB1; chromosome 4q24) and the p65 (RELA; 11q13). NFκB is maintained in the cytosol in an inactive state, bound with members of the IκB family that inhibit nuclear transport or block NFκB's nuclear translocation signal [137]. Activation usually proceeds by the IKK kinase complex phosphorylating IκB, resulting in IκB ubiquitination and degradation [138]. NFκB (RELA/NFκB1) is implicated in several critical cellular functions [139]. Reflecting its regulation by both estrogen and growth factors [140,141] that are involved in endocrine resistance [10,142], normal mammary gland development is dependent upon NFκB [143]. Increased NFκB activity arises during neoplastic transformation in the rat [144] and mouse mammary gland [145]. Upregulation of NFκB is associated with E2-independence [140,143]. The predominant NFκB form in breast cancer cell lines is RELA/NFκB1; the p52 family member also is expressed in some breast cancers [146].

We have shown that NFκB can confer estrogen-independence and AE crossresistance [75,95,147]. Estrogen-independent growth *in vitro* and *in vivo* is supported by increases in both NFκB DNA binding activity and expression of BCL3 [147]. This study highlights the functional implications of NFκB in AI resistance. Expression of IκBα (NFκB repressor) in estrogen-independent LCC1 cells (LCC1 cells are derived from MCF-7 and are estrogen-independent but sensitive to AEs [148]), which have increased NFκB activation relative to estrogen-dependent MCF-7 cells, eliminates their estrogen-independence *in vivo*.

LCC9 cells (TAM and FAS crossresistant variant of LCC1 [28]) exhibit a further increase in NFκB expression and activation relative to LCC1 cells, apparently driven by increased expression of NEMO (IKKγ) [75]. These observations imply that the level of activity in LCC1 cells is adequate for estrogen-independence but not AE resistance. Increased activation of NFκB [95] and loss of its antiestrogenic regulation in LCC9 cells [75] suggest that these cells might be dependent upon NFκB for survival/growth. Thus, we compared the growth response of LCC1 and LCC9 cells to vehicle or parthenolide (300 and 600 nM), a small molecule inhibitor of NFκB [149]. Parthenolide produces a dose-dependent inhibition of MCF7/LCC9 cells with an apparent IC_{50} of approximately 600 nM ($p < 0.01$ at both 300 and 600 nM parthenolide). In marked contrast, parthenolide does not affect growth of LCC1 cells at either of these concentrations [75]. We next asked if parthenolide can re-sensitize LCC9 cells to FAS-mediated apoptosis. FAS and parthenolide synergize to induce LCC9 cell death [75]. Since FAS alone is inactive [28], this synergism reflects at least a partial reversal of the FAS resistance component of the LCC9 cell phenotype and implicates NFκB as a key determinant [75]. Thus, AE crossresistant cells exhibit a greater reliance upon NFκB signaling for proliferation, and inhibition of NFκB restores their sensitivity to apoptosis induced by FAS [95].

4.4. Expression of ER, PGR, XBP1, NFκB and IRF1 in breast tumors

Using gene expression microarrays, we previously compared the global structures of the transcriptomes of three ER+ human breast cancer cell lines (MCF-7, T47D, ZR75-1) and 13 human breast tumors (11 ER+; 2 ER−) and showed these to be notably similar to ER+ breast tumors from patients [150]. The striking similarities between cell lines and tumors are supported by a report that the estrogen-regulated genes in these cell lines are similarly regulated in breast tumors [151]. These data show that ER+ breast cancer cell lines and ER+ breast tumors in women share global similarities in the structures of their respective transcriptomes [150], and that these cell lines are appropriate models in which to identify clinically relevant endocrine-regulated molecular events [150,151]. Nonetheless, it is necessary to show that the seed genes we have selected are likely to be relevant to the biology of ER+ breast tumors.

To begin to explore the possible clinical relevance of these functional studies, we first asked if we could detect XBP1, NFκB, and IRF1 in breast tumors. We then asked whether any of these proteins were coexpressed in patterns consistent with the experimental data from cell lines. Using a series of breast cancer tissue arrays comprising 480 cores from 54 breast carcinomas (mostly ER+ tumors), we applied immunohistochemistry to explore the expression of the seed genes [152]. Pairwise correlation analyses cannot account for the possibility that unknown associations among proteins may confound each other, so we applied a novel use of partial correlation coefficient analysis. Partial correlation analysis allows an estimate of the correlation between two variables while controlling for a third, fourth and/or fifth and is particularly useful in the analysis of small signaling networks of 3–5 variables [153].

We confirmed the well established coexpression of ERα and PgR, implying that the samples are representative of most ER+ breast cancers. XBP1, NFκB, and IRF1 are each found in a high proportion of breast tumors [152]. Total XBP1 was measured, as XBP1(S) antibodies were not then available. XBP1 staining is variable but detectable in 79% of breast tumors. A very recent study has reported a significant association between XBP1(S) mRNA and poor response to endocrine therapy [154]—entirely consistent with our studies in breast cancer cell lines [76]. 57% of the tumors express detectable RELA in their neoplastic cells, similar to a prior study of $n = 17$ breast tumors [146].

Expression of several of the proteins is correlated in breast tumors. IRF1 correlates with ER and PGR, and also with RELA and XBP1. While, these correlations depend on the subcellular localization of IRF1 and some are direct and others inverse correlations, they are fully consistent with the interpretation that these expression patterns reflect functionally relevant signaling links. For example, we might predict that IRF1 sequestered in the cytosol, unlike that in the nucleus, cannot act as a proapoptotic TF (the full coexpression patterns are described detail in the report by Zhu et al. [152]). We also find coexpression of XBP1 and RELA, consistent with the observation that XBP1 may be downstream of NFκB [109]. When each of the significant correlations is examined in the partial correlation coefficient models, the IRF1, NFκB, and XBP correlations remain [152]. These data are consistent with these three reflecting some component of a larger signaling network active in some ER+ breast cancers and further support their selection as seed genes from which to grow this network and understand its topology and function. Moreover, the functional data from our experimental models implies that this network links signaling and function through two key subcellular components—mitochondria and the endoplasmic reticulum.

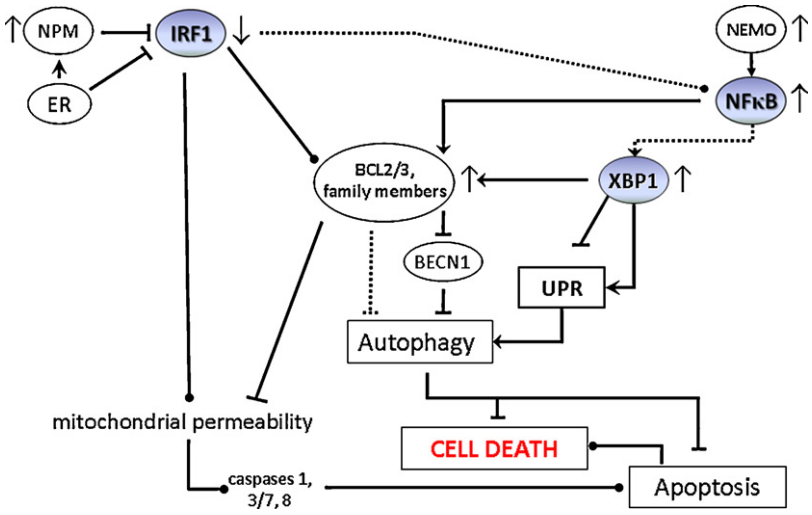


Fig. 5. Endocrine resistance seed-gene network. Simple representation of a seed-gene network of XBP1, NFκB and IRF1 based on functional data obtained from an appropriate cellular context (resistant MCF7/LCC9 cells).

4.5. Simple representation of a seed-gene network of XBP1, NFκB and IRF1 based on functional data obtained from an appropriate cellular context

The experimental data supporting the wiring-diagram representation of the network model shown in Fig. 5 are discussed the preceding sections. Here we discuss how the signals may flow through this network. The three primary seed genes of IRF1, XBP1, and NFκB are evident as previously proposed [95]. IRF1 expression is repressed in resistant cells [95] but induced by antiestrogens in sensitive cells [65]. A dominant negative IRF1 confers an antiestrogen resistant phenotype, implying that IRF1-driven prodeath signaling is key to the regulation of cell fate [65].

In addition to changes in the expression of IRF1, the upregulation of NPM expression [95,155] could also affect IRF1 action. Both NPM and IRF1 are estrogen-regulated genes in MCF-7 cells, IRF1 expression being suppressed, whereas NPM is induced [129,155]. Since NPM inhibits the transcription regulatory activities of IRF1 [156], the increase in NPM expression could bind remaining IRF1 and inhibit its ability to initiate an apoptotic caspase cascade. We also cannot exclude the possibility that NPM has activities independent of blocking IRF1, since NPM overexpression is sufficient to transform NIH 3T3 cells in a standard oncogenesis assay [156]. Increased levels of serum autoantibodies to NPM predict recurrence on TAM 6-months prior to clinical detection [157].

IRF1 and NFκB are known to form heterodimers and to regulate directly gene expression [158,159] including that of the inducible nitric oxide synthase promoter [158]. Since we do not know if it is primarily the gene regulatory effects of these heterodimers, or if their subcellular location is key (they act by preferentially sequestering one or the other so that transcriptional regulation does not occur), this is shown as a dotted line. We would predict, based on the inverse expression between NFκB and IRF1 in LCC9 cells [95] and in some breast cancers [152], that either the prodeath effects of any remaining IRF1 are being sequestered by NFκB in resistant cells and/or that the overexpression and activation of NFκB leads to a dominance of its prosurvival activities. The increased sensitivity of resistant cells to parthenolide is consistent with the functional relevance of at least the latter signaling outcome [75].

We have previously shown that the upregulation of NFκB in antiestrogen resistant cells [95] is likely driven in part by increased NEMO/IKKγ activity [75]. The prosurvival activities of NFκB are well documented [160]. Precisely how NFκB regulates cell survival remains to be fully established but activation of prosurvival members of the BCL2 gene family are involved in both acquired estrogen-independence [147] and antiestrogen resistance [75,76]. While NFκB is predicted to induce transcription of XBP1 [109], we have yet to report this direct regulation in breast cancer cells (studies are in progress). Whether or not this occurs, XBP1 is clearly upregulated in resistant cells [95] and this activity is sufficient to confer both estrogen-independence and antiestrogen resistance [76]. More recently, increased XBP1 mRNA expression has been show to predict for a poor response to TAM in breast cancer patients [154].

The central role of XBP1 within the UPR clearly implicates UPR activation in responsiveness to both estrogen-withdrawal and antiestrogen treatment [76]. UPR also is known to induce autophagy [102], although whether this is a prosurvival or prodeath autophagy remains unclear in the context of determining endocrine responsiveness. Autophagy is regulated, at least in part, by the action of BECN1. BECN1 activity is regulated by BCL2, which binds BECN1 and can block BECN1-mediated autophagy [36].

The regulation of BCL2 family members (BCL2, BCL3, and probably others) whether by IRF1, NFκB, and/or XBP1, can affect both autophagy and the intrinsic apoptosis pathway. The intersection of their signaling at BCL2 family members, as shown in Fig. 5, is one location within the broader network where the balance between prodeath and prosurvival signaling, and whether prodeath is autophagic or apoptotic, is determined. This intersection also links signaling through the UPR and endoplasmic reticulum to the mitochondria with the cell fate decision mechanisms—at least in the context of determining cell fate in the context of endocrine responsiveness in breast cancer. The signaling depicted in Fig. 5 represents only a small component of this broader network. Nevertheless, this initial wiring-diagram is consistent with a body of functional data in experimental models and it provides sufficient seed genes, their edges, and the directionality of these edges, to begin a more detailed exploration of this central network. Understanding this network's topology and function will lead to better candidates for drug discovery and to better algorithms to predict how individual tumors will respond to specific endocrine therapies.

References

[1] A. Jemal, R. Siegel, E. Ward, T. Murray, J. Xu, C. Smigal, M.J. Thun, Cancer statistics, 2006, CA Cancer J. Clin. 56 (2006) 106–130.
[2] W.D. Foulkes, K. Metcalfe, P. Sun, W.M. Hanna, H.T. Lynch, P. Ghadirian, N. Tung, O.I. Olopade, B.L. Weber, J. McLennan, I.A. Olivotto, L.R. Begin, S.A. Narod, Estrogen receptor status in BRCA1- and BRCA2-related breast cancer: the influ-

ence of age, grade, and histological type, Clin. Cancer Res. 10 (2004) 2029–2034.
[3] S.M. Thorpe, Estrogen and progesterone receptor determinations in breast cancer. Technology, biology and clinical significance, Acta Oncol. 27 (1988) 1–19.
[4] W.D. Foulkes, Re: Estrogen receptor status of primary breast cancer is predictive of estrogen receptor status of contralateral breast cancer, J. Natl. Cancer Inst. 96 (2004) 1040–1041.
[5] Early Breast Cancer Trialists' Collaborative Group, Polychemotherapy for early breast cancer: an overview of randomised trials, Lancet 352 (1998) 930–942.
[6] B. Fisher, J. Dignam, E.P. Mamounas, J.P. Costantino, D.L. Wickerham, C. Redmond, N. Wolmark, N.V. Dimitrov, D.M. Bowman, A.G. Glass, J.L. Atkins, N. Abramson, C.M. Sutherland, B.S. Aron, R.G. Margolese, Sequential methotrexate and fluorouracil for the treatment of node-negative breast cancer patients with estrogen-receptor-negative tumors: eight year results from National Surgical Adjuvant Breast and Bowel Project (NSABP) B-13 and first report of findings from NSABP B-19 comparing methotrexate and fluorouracil with conventional cyclophosphamide, methotrexate, and fluorouracil, J. Clin. Oncol. 14 (1996) 1982–1992.
[7] E.G. Mansour, R. Gray, A.H. Shatila, D.C. Tormey, M.R. Cooper, C.K. Osborne, G. Falkson, Survival advantage of adjuvant chemotherapy in high-risk node-negative breast cancer: ten-year analysis—an intergroup study, J. Clin. Oncol. 16 (1998) 3486–3492.
[8] Early Breast Cancer Trialists' Collaborative Group, Tamoxifen for early breast cancer: an overview of the randomized trials, Lancet 351 (1998) 1451–1467.
[9] Early Breast Cancer Trialists Collaborative Group, Systemic treatment of early breast cancer by hormonal, cytotoxic, or immune therapy, Lancet 399 (1992) 1–15.
[10] R. Clarke, F. Leonessa, J.N. Welch, T.C. Skaar, Cellular and molecular pharmacology of antiestrogen action and resistance, Pharmacol. Rev. 53 (2001) 25–71.
[11] M.P. Cole, C.T.A. Jones, I.D.H. Todd, A new antioestrogenic agent in late breast cancer. An early clinical appraisal of ICI 46474, Br. J. Cancer 25 (1971) 270–275.
[12] S.R. Cummings, S. Eckert, K.A. Krueger, D. Grady, T.J. Powles, J.A. Cauley, L. Norton, T. Nickelsen, N.H. Bjarnson, M. Morrow, M.E. Lippman, D. Black, J.E. Glusman, A. Costa, V.C. Jordan, The effect of raloxifene on risk of breast cancer in postmenopausal women; results from the MORE randomized trial. Multiple Outcomes of Raloxifene Evaluation, J. Am. Med. Assoc. 281 (1999) 2189–2197.
[13] A. Howell, D. DeFriend, J.F.R. Robertson, R.W. Blamey, P. Walton, Response to a specific antioestrogen (ICI 182,780) in tamoxifen-resistant breast cancer, Lancet 345 (1995) 29–30.
[14] A. Howell, J.F. Robertson, A.J. Quaresma, A. Aschermannova, L. Mauriac, U.R. Kleeberg, I. Vergote, B. Erikstein, A. Webster, C. Morris, Fulvestrant, formerly ICI 182, 780, is as effective as anastrozole in postmenopausal women with advanced breast cancer progressing after prior endocrine treatment, J. Clin. Oncol. 20 (2002) 3396–3403.
[15] G. Ferretti, E. Bria, D. Giannarelli, A. Felici, P. Papaldo, A. Fabi, C.S. Di, E.M. Ruggeri, M. Milella, M. Ciccarese, F.L. Cecere, A. Gelibter, C. Nuzzo, F. Cognetti, E. Terzoli, P. Carlini, Second- and third-generation aromatase inhibitors as first-line endocrine therapy in postmenopausal metastatic breast cancer patients: a pooled analysis of the randomised trials, Br. J. Cancer 94 (2006) 1789–1796.
[16] B. Thurlimann, A. Keshaviah, A.S. Coates, H. Mouridsen, L. Mauriac, J.F. Forbes, R. Paridaens, M. Castiglione-Gertsch, R.D. Gelber, M. Rabaglio, I. Smith, A. Wardley, K.N. Price, A. Goldhirsch, A comparison of letrozole and tamoxifen in postmenopausal women with early breast cancer, N. Engl. J. Med. 353 (2005) 2747–2757.
[17] R.C. Coombes, E. Hall, L.J. Gibson, R. Paridaens, J. Jassem, T. Delozier, S.E. Jones, I. Alvarez, G. Bertelli, O. Ortmann, A.S. Coates, E. Bajetta, D. Dodwell, R.E. Coleman, L.J. Fallowfield, E. Mickiewicz, J. Andersen, P.E. Lonning, G. Cocconi, A. Stewart, N. Stuart, C.F. Snowdon, M. Carpentieri, G. Massimini, J.M. Bliss, A randomized trial of exemestane after two to three years of tamoxifen therapy in postmenopausal women with primary breast cancer, N. Engl. J. Med. 350 (2004) 1081–1092.
[18] F. Boccardo, A. Rubagotti, D. Amoroso, M. Mesiti, D. Romeo, C. Caroti, A. Farris, G. Cruciani, E. Villa, G. Schieppati, G. Mustacchi, Sequential tamoxifen and aminoglutethimide versus tamoxifen alone in the adjuvant treatment of postmenopausal breast cancer patients: results of an Italian cooperative study, J. Clin. Oncol. 19 (2001) 4209–4215.
[19] R. Jakesz, W. Jonat, M. Gnant, M. Mittlboeck, R. Greil, C. Tausch, J. Hilfrich, W. Kwasny, C. Menzel, H. Samonigg, M. Seifert, G. Gademann, M. Kaufmann, J. Wolfgang, Switching of postmenopausal women with endocrine-responsive early breast cancer to anastrozole after 2 years' adjuvant tamoxifen: combined results of ABCSG trial 8 and ARNO 95 trial, Lancet 366 (2005) 455–462.
[20] M. Baum, A. Buzdar, J. Cuzick, J. Forbes, J. Houghton, A. Howell, T. Sahmoud, Anastrozole alone or in combination with tamoxifen versus tamoxifen alone for adjuvant treatment of postmenopausal women with early-stage breast cancer: results of the ATAC (Arimidex, Tamoxifen Alone or in Combination) trial efficacy and safety update analyses, Cancer 98 (2003) 1802–1810.
[21] J. Bonneterre, A. Buzdar, J.M. Nabholtz, J.F. Robertson, B. Thurlimann, M. von Euler, T. Sahmoud, A. Webster, M. Steinberg, Anastrozole is superior to tamoxifen as first-line therapy in hormone receptor positive advanced breast carcinoma, Cancer 92 (2001) 2247–2258.
[22] H. Mouridsen, M. Gershanovich, Y. Sun, R. Perez-Carrion, C. Boni, A. Monnier, J. Apffelstaedt, R. Smith, H.P. Sleeboom, F. Janicke, A. Pluzanska, M. Dank, D. Becquart, P.P. Bapsy, E. Salminen, R. Snyder, M. Lassus, J.A. Verbeek, B. Staffler, H.A. Chaudri-Ross, M. Dugan, Superior efficacy of letrozole versus tamoxifen as first-line therapy for postmenopausal women with advanced breast cancer: results of a phase iii study of the international letrozole breast cancer group, J. Clin. Oncol. 19 (2001) 2596–2606.
[23] R. Clarke, N. Brünner, Cross resistance and molecular mechanisms in antiestrogen resistance, Endocr. Relat. Cancer 2 (1995) 59–72.
[24] C. Osipo, C. Gajdos, H. Liu, B. Chen, V.C. Jordan, Paradoxical action of fulvestrant in estradiol-induced regression of tamoxifen-stimulated breast cancer, J. Natl. Cancer Inst. 95 (2003) 1597–1608.
[25] M.M. Gottardis, V.C. Jordan, Development of tamoxifen-stimulated growth of MCF-7 tumors in athymic mice after long-term antiestrogen administration, Cancer Res. 48 (1988) 5183–5187.
[26] C.K. Osborne, E.B. Coronado, J.P. Robinson, Human breast cancer in athymic nude mice: cytostatic effects of long-term antiestrogen therapy, Eur. J. Cancer Clin. Oncol. 23 (1987) 1189–1196.
[27] N. Brünner, T.L. Frandsen, C. Holst-Hansen, M. Bei, E.W. Thompson, A.E. Wakeling, M.E. Lippman, R. Clarke, MCF7/LCC2: A 4-hydroxytamoxifen resistant human breast cancer variant which retains sensitivity to the steroidal antiestrogen ICI 182, 780, Cancer Res. 53 (1993) 3229–3232.
[28] N. Brünner, B. Boysen, S. Jirus, T.C. Skaar, C. Holst-Hansen, J. Lippman, T. Frandsen, M. Spang-Thomsen, S.A.W. Fuqua, R. Clarke, MCF7/LCC9: an antiestrogen resistant MCF-7 variant in which acquired resistance to the steroidal antiestrogen ICI 182, 780 confers an early crossresistance to the non-steroidal antiestrogen tamoxifen, Cancer Res. 57 (1997) 3486–3493.
[29] R. Clarke, M.C. Liu, K.B. Bouker, Z. Gu, R.Y. Lee, Y. Zhu, T.C. Skaar, B. Gomez, K. O'Brien, Y. Wang, L.A. Hilakivi-Clarke, Antiestrogen resistance in breast cancer and the role of estrogen receptor signaling, Oncogene 22 (2003) 7316–7339.
[30] Y. Nomura, H. Tashiro, K. Hisamatsu, Differential effects of estrogen and antiestrogen on in vitro clonogenic growth of human breast cancers in soft agar, J. Natl. Cancer Inst. 82 (1990) 1146–1149.
[31] A. Haddow, J.M. Watkins, E. Paterson, Influence of synthetic oestrogens upon advanced malignant disease, Br. Med. J. 2 (1944) 393–398.
[32] A.L. Walpole, E. Paterson, Synthetic oestrogens in mammary cancer, Lancet 2 (1949) 783–789.
[33] R. Clarke, H.W. van den Berg, R.F. Murphy, Tamoxifen and 17β-estradiol reduce the membrane fluidity of human breast cancer cells, J. Natl. Cancer Inst. 82 (1990) 1702–1705.
[34] S. Pattingre, B. Levine, Bcl-2 inhibition of autophagy: a new route to cancer? Cancer Res. 66 (2006) 2885–2888.
[35] R. Riggins, A.H. Bouton, M.C. Liu, R. Clarke, Antiestrogens, aromatase inhibitors, and apoptosis in breast cancer, Vitam. Horm. 71 (2005) 201–237.
[36] S. Pattingre, A. Tassa, X. Qu, R. Garuti, X.H. Liang, N. Mizushima, M. Packer, M.D. Schneider, B. Levine, Bcl-2 antiapoptotic proteins inhibit Beclin 1-dependent autophagy, Cell 122 (2005) 927–939.
[37] B. Zhivotovsky, G. Kroemer, Apoptosis and genomic instability, Nat. Rev. Mol. Cell Biol. 5 (2004) 752–762.
[38] J.M. Brown, L.D. Attardi, The role of apoptosis in cancer development and treatment response, Nat. Rev. Cancer 5 (2005) 231–237.
[39] J.C. Goldstein, N.J. Waterhouse, P. Juin, G.I. Evan, D.R. Green, The coordinate release of cytochrome c during apoptosis is rapid, complete and kinetically invariant, Nat. Cell Biol. 2 (2000) 156–162.
[40] S. Shimizu, T. Kanaseki, N. Mizushima, T. Mizuta, S. Rakawa-Kobayashi, C.B. Thompson, Y. Tsujimoto, Role of Bcl-2 family proteins in a non-apoptotic programmed cell death dependent on autophagy genes, Nat. Cell Biol. 6 (2004) 1221–1228.
[41] M. Lamparska-Przybysz, B. Gajkowska, T. Motyl, BID-deficient breast cancer MCF-7 cells as a model for the study of autophagy in cancer therapy, Autophagy 2 (2006) 47–48.
[42] N. Furuya, J. Yu, M. Byfield, S. Pattingre, B. Levine, The evolutionarily conserved domain of Beclin 1 is required for Vps34 binding, autophagy and tumor suppressor function, Autophagy 1 (2005) 46–52.
[43] X.H. Liang, S. Jackson, M. Seaman, K. Brown, B. Kempkes, H. Hibshoosh, B. Levine, Induction of autophagy and inhibition of tumorigenesis by beclin 1, Nature 402 (1999) 672–676.
[44] A.L. Edinger, C.B. Thompson, Death by design: apoptosis, necrosis and autophagy, Curr. Opin. Cell Biol. 16 (2004) 663–669.
[45] L.E. Broker, F.A. Kruyt, G. Giaccone, Cell death independent of caspases: a review, Clin. Cancer Res. 11 (2005) 3155–3162.
[46] N. Mizushima, T. Yoshimori, How to interpret LC3 immunoblotting, Autophagy 3 (2007) 542–545.
[47] S. Kimura, T. Noda, T. Yoshimori, Dissection of the autophagosome maturation process by a novel reporter protein, tandem fluorescent-tagged LC3, Autophagy 3 (2007) 452–460.
[48] D.J. Klionsky, H. Abeliovich, P. Agostinis, D.K. Agrawal, G. Aliev, D.S. Askew, M. Baba, E.H. Baehrecke, B.A. Bahr, A. Ballabio, B.A. Bamber, D.C. Bassham, E. Bergamini, X. Bi, M. Biard-Piechaczyk, J.S. Blum, D.E. Bredesen, J.L. Brodsky, J.H. Brumell, U.T. Brunk, W. Bursch, N. Camougrand, E. Cebollero, F. Cecconi, Y. Chen, L.S. Chin, A. Choi, C.T. Chu, J. Chung, P.G. Clarke, R.S. Clark, S.G. Clarke, C. Clave, J.L. Cleveland, P. Codogno, M.I. Colombo, A. Coto-Montes, J.M. Cregg, A.M. Cuervo, J. Debnath, F. Demarchi, P.B. Dennis, P.A. Dennis, V. Deretic, R.J. Devenish, S.F. Di, J.F. Dice, M. Difiglia, S. Nesh-Kumar, C.W. Distelhorst, M. Djavaheri-Mergny, F.C. Dorsey, W. Droge, M. Dron, W.A. Dunn Jr., M. Duszenko, N.T. Eissa, Z. Elazar, A. Esclatine, E.L. Eskelinen, L. Fesus, K.D. Finley, J.M. Fuentes, J. Fueyo, K. Fujisaki, B. Galliot, F.B. Gao, D.A. Gewirtz, S.B. Gibson, A. Gohla, A.L. Goldberg, R. Gonzalez, C. Gonzalez-Estevez, S. Gorski, R.A. Gottlieb, D. Haussinger, Y.W. He, K. Heidenreich, J.A. Hill, M. Hoyer-Hansen, X. Hu, W.P. Huang, A. Iwasaki, M. Jaattela, W.T. Jackson, X. Jiang, S. Jin, T. Johansen, J.U.

Jung, M. Kadowaki, C. Kang, A. Kelekar, D.H. Kessel, J.A. Kiel, H.P. Kim, A. Kimchi, T.J. Kinsella, K. Kiselyov, K. Kitamoto, E. Knecht, M. Komatsu, E. Kominami, S. Kondo, A.L. Kovacs, G. Kroemer, C.Y. Kuan, R. Kumar, M. Kundu, J. Landry, M. Laporte, W. Le, H.Y. Lei, M.J. Lenardo, B. Levine, A. Lieberman, K.L. Lim, F.C. Lin, W. Liou, L.F. Liu, G. Lopez-Berestein, C. Lopez-Otin, B. Lu, K.F. Macleod, W. Malorni, W. Martinet, K. Matsuoka, J. Mautner, A.J. Meijer, A. Melendez, P. Michels, G. Miotto, W.P. Mistiaen, N. Mizushima, B. Mograbi, I. Monastyrska, M.N. Moore, P.I. Moreira, Y. Moriyasu, T. Motyl, C. Munz, L.O. Murphy, N.I. Naqvi, T.P. Neufeld, I. Nishino, R.A. Nixon, T. Noda, B. Nurnberg, M. Ogawa, N.L. Oleinick, L.J. Olsen, B. Ozpolat, S. Paglin, G.E. Palmer, I. Papassideri, M. Parkes, D.H. Perlmutter, G. Perry, M. Piacentini, R. Pinkas-Kramarski, M. Prescott, T. Proikas-Cezanne, N. Raben, A. Rami, F. Reggiori, B. Rohrer, D.C. Rubinsztein, K.M. Ryan, J. Sadoshima, H. Sakagami, Y. Sakai, M. Sandri, C. Sasakawa, M. Sass, C. Schneider, P.O. Seglen, O. Seleverstov, J. Settleman, J.J. Shacka, I.M. Shapiro, A. Sibirny, E.C. Silva-Zacarin, H.U. Simon, C. Simone, A. Simonsen, M.A. Smith, K. Spanel-Borowski, V. Srinivas, M. Steeves, H. Stenmark, P.E. Stromhaug, C.S. Subauste, S. Sugimoto, D. Sulzer, T. Suzuki, M.S. Swanson, I. Tabas, F. Takeshita, N.J. Talbot, Z. Talloczy, K. Tanaka, K. Tanaka, I. Tanida, G.S. Taylor, J.P. Taylor, A. Terman, G. Tettamanti, C.B. Thompson, M. Thumm, A.M. Tolkovsky, S.A. Tooze, R. Truant, L.V. Tumanovska, Y. Uchiyama, T. Ueno, N.L. Uzcategui, d.K. van I, E.C. Vaquero, T. Vellai, M.W. Vogel, H.G. Wang, P. Webster, J.W. Wiley, Z. Xi, G. Xiao, J. Yahalom, J.M. Yang, G. Yap, X.M. Yin, T. Yoshimori, L. Yu, Z. Yue, M. Yuzaki, O. Zabirnyk, X. Zheng, X. Zhu, R.L. Deter, Guidelines for the use and interpretation of assays for monitoring autophagy in higher eukaryotes, Autophagy 4 (2008) 151–175.

[49] S. Yousefi, R. Perozzo, I. Schmid, A. Ziemiecki, T. Schaffner, L. Scapozza, T. Brunner, H.U. Simon, Calpain-mediated cleavage of Atg5 switches autophagy to apoptosis, Nat. Cell Biol. 8 (2006) 1124–1132.

[50] S. Jin, Autophagy, mitochondrial quality control, and oncogenesis, Autophagy 2 (2006) 80–84.

[51] M.A. Park, A. Yacoub, D. Sarkar, L. Emdad, M. Rahmani, S. Spiegel, C. Koumenis, M. Graf, D.T. Curiel, S. Grant, P.B. Fisher, P. Dent, PERK-dependent regulation of MDA-7/IL-24-induced autophagy in primary human glioma cells, Autophagy 4 (2008).

[52] A. Yacoub, M.A. Park, P. Gupta, M. Rahmani, G. Zhang, H. Hamed, D. Hanna, D. Sarkar, I.V. Lebedeva, L. Emdad, M. Sauane, N. Vozhilla, S. Spiegel, C. Koumenis, M. Graf, D.T. Curiel, S. Grant, P.B. Fisher, P. Dent, Caspase-, cathepsin-, and PERK-dependent regulation of MDA-7/IL-24-induced cell killing in primary human glioma cells, Mol. Cancer Ther. 7 (2008) 297–313.

[53] Y. Chen, E. Millan-Ward, J. Kong, S.J. Israels, S.B. Gibson, Oxidative stress induces autophagic cell death independent of apoptosis in transformed and cancer cells, Cell Death Differ. 15 (2008) 171–182.

[54] C.H. Yan, Y.P. Yang, Z.H. Qin, Z.L. Gu, P. Reid, Z.Q. Liang, Autophagy is involved in cytotoxic effects of crotoxin in human breast cancer cell line MCF-7 cells, Acta Pharmacol. Sin. 28 (2007) 540–548.

[55] A. Guillon-Munos, M.X. van Bemmelen, P.G. Clarke, Autophagy can be a killer even in apoptosis-competent cells, Autophagy 2 (2006) 140–142.

[56] A. Apel, I. Herr, H. Schwarz, H.P. Rodemann, A. Mayer, Blocked autophagy sensitizes resistant carcinoma cells to radiation therapy, Cancer Res. 68 (2008) 1485–1494.

[57] M. Li, X. Jiang, D. Liu, Y. Na, G.F. Gao, Z. Xi, Autophagy protects LNCaP cells under androgen deprivation conditions, Autophagy 4 (2008) 54–60.

[58] P. Boya, R.A. Gonzalez-Polo, N. Casares, J.L. Perfettini, P. Dessen, N. Larochette, D. Metivier, D. Meley, S. Souquere, T. Yoshimori, G. Pierron, P. Codogno, G. Kroemer, Inhibition of macroautophagy triggers apoptosis, Mol. Cell Biol. 25 (2005) 1025–1040.

[59] M. Castedo, J.L. Perfettini, T. Roumier, A. Valent, H. Raslova, K. Yakushijin, D. Horne, J. Feunteun, G. Lenoir, R. Medema, W. Vainchenker, G. Kroemer, Mitotic catastrophe constitutes a special case of apoptosis whose suppression entails aneuploidy, Oncogene 23 (2004) 4362–4370.

[60] I.B. Roninson, E.V. Broude, B.D. Chang, If not apoptosis, then what? Treatment-induced senescence and mitotic catastrophe in tumor cells, Drug Resist. Updat. 4 (2001) 303–313.

[61] G.J. Kops, B.A. Weaver, D.W. Cleveland, On the road to cancer: aneuploidy and the mitotic checkpoint, Nat. Rev. Cancer 5 (2005) 773–785.

[62] D. Kanduc, A. Mittelman, R. Serpico, E. Sinigaglia, A.A. Sinha, C. Natale, R. Santacroce, M.G. Di Corcia, A. Lucchese, L. Dini, P. Pani, S. Santacroce, S. Simone, R. Bucci, E. Farber, Cell death: apoptosis versus necrosis, Int. J. Oncol. 21 (2002) 165–170.

[63] G.P. Dimri, X. Lee, G. Basile, M. Acosta, G. Scott, C. Roskelley, E.E. Medrano, M. Linskens, I. Rubelj, O. Pereira-Smith, A biomarker that identifies senescent human cells in culture and in aging skin in vivo, Proc. Natl. Acad. Sci. U.S.A. 92 (1995) 9363–9367.

[64] W.E. Wright, O.M. Pereira-Smith, J.W. Shay, Reversible cellular senescence: implications for immortalization of normal human diploid fibroblasts, Mol. Cell Biol. 9 (1989) 3088–3092.

[65] K.B. Bouker, T.C. Skaar, D.R. Fernandez, K.A. O'Brien, R. Clarke, Interferon regulatory factor-1 mediates the proapoptotic but not cell cycle arrest effects of the steroidal antiestrogen ICI 182, 780 (Faslodex, Fulvestrant), Cancer Res. 64 (2004) 4030–4039.

[66] N. Kyprianou, H.F. English, N.E. Davidson, J.T. Isaacs, Programmed cell death during regression of the MCF-7 human breast cancer following estrogen ablation, Cancer Res. 51 (1991) 162–166.

[67] M.F. El Etreby, Y. Liang, R.W. Wrenn, P.V. Schoenlein, Additive effect of mifepristone and tamoxifen on apoptotic pathways in MCF-7 human breast cancer cells, Breast Cancer Res. Treat. 51 (1998) 149–168.

[68] V.T. Gaddy, J.T. Barrett, J.N. Delk, A.M. Kallab, A.G. Porter, P.V. Schoenlein, Mifepristone induces growth arrest, caspase activation, and apoptosis of estrogen receptor-expressing, antiestrogen-resistant breast cancer cells, Clin. Cancer Res. 10 (2004) 5215–5225.

[69] W. Bursch, A. Ellinger, H. Kienzl, L. Torok, S. Pandey, M. Sikorska, R. Walker, R.S. Hermann, Active cell death induced by the anti-estrogens tamoxifen and ICI 164 384 in human mammary carcinoma cells (MCF-7) in culture: the role of autophagy, Carcinogenesis 17 (1996) 1595–1607.

[70] B. Inbal, S. Bialik, I. Sabanay, G. Shani, A. Kimchi, DAP kinase and DRP-1 mediate membrane blebbing and the formation of autophagic vesicles during programmed cell death, J. Cell Biol. 157 (2002) 455–468.

[71] M.A. Qadir, B. Kwok, W.H. Dragowska, K.H. To, D. Le, M.B. Bally, S.M. Gorski, Macroautophagy inhibition sensitizes tamoxifen-resistant breast cancer cells and enhances mitochondrial depolarization, Breast Cancer Res. Treat. 3 (2008) 389–403.

[72] G. Bjorkoy, T. Lamark, A. Brech, H. Outzen, M. Perander, A. Overvatn, H. Stenmark, T. Johansen, p62/SQSTM1 forms protein aggregates degraded by autophagy and has a protective effect on huntingtin-induced cell death, J. Cell Biol. 171 (2005) 603–614.

[73] Y. Ichimura, T. Kumanomidou, Y.S. Sou, T. Mizushima, J. Ezaki, T. Ueno, E. Kominami, T. Yamane, K. Tanaka, M. Komatsu, Structural basis for sorting mechanism of p62 in selective autophagy, J. Biol. Chem. 283 (2008) 22847–22857.

[74] Y. Ichimura, E. Kominami, K. Tanaka, M. Komatsu, Selective turnover of p62/A170/SQSTM1 by autophagy, Autophagy 4 (2008).

[75] R. Riggins, A. Zwart, N. Nehra, P. Agarwal, R. Clarke, The NFκB inhibitor parthenolide restores ICI 182, 780 (Faslodex; Fulvestrant)-induced apoptosis in antiestrogen resistant breast cancer cells, Mol. Cancer Ther. 4 (2005) 33–41.

[76] B.P. Gomez, R. Riggins, A.N. Shajahan, U. Klimach, A. Wang, A.C. Crawford, Y. Zhu, A. Zwart, M. Wang, R. Clarke, Human X-Box binding protein-1 confers both estrogen independence and antiestrogen resistance in breast cancer cell lines, FASEB J. 21 (2007) 4013–4027.

[77] R. Clarke, N. Brünner, B.S. Katzenellenbogen, E.W. Thompson, M.J. Norman, C. Koppi, S. Paik, M.E. Lippman, R.B. Dickson, Progression from hormone dependent to hormone independent growth in MCF-7 human breast cancer cells, Proc. Natl. Acad. Sci. U.S.A. 86 (1989) 3649–3653.

[78] R. Clarke, N. Brünner, E.W. Thompson, P. Glanz, D. Katz, R.B. Dickson, M.E. Lippman, The inter-relationships between ovarian-independent growth, antiestrogen resistance and invasiveness in the malignant progression of human breast cancer, J. Endocrinol. 122 (1989) 331–340.

[79] D.A. Cameron, J.C. Keen, J.M. Dixon, C. Bellamy, A. Hanby, T.J. Anderson, W.R. Miller, Effective tamoxifen therapy of breast cancer involves both antiproliferative and pro-apoptotic changes, Eur. J. Cancer 36 (2000) 845–851.

[80] T. Reimer, D. Koczan, H. Muller, K. Friese, H.J. Thiesen, B. Gerber, Tumour Fas ligand:Fas ratio greater than 1 is an independent marker of relative resistance to tamoxifen therapy in hormone receptor positive breast cancer, Breast Cancer Res. 4 (2002) R9.

[81] P.A. Ellis, I.E. Smith, S. Detre, S.A. Burton, J. Salter, R. A'Hern, G. Walsh, S.R. Johnston, M. Dowsett, Reduced apoptosis and proliferation and increased Bcl-2 in residual breast cancer following preoperative chemotherapy, Breast Cancer Res. Treat. 48 (1998) 107–116.

[82] N. Brunner, V. Boulay, A. Fojo, C.E. Freter, M.E. Lippman, R. Clarke, Acquisition of hormone-independent growth in MCF-7 cells is accompanied by increased expression of estrogen-regulated genes but without detectable DNA amplifications, Cancer Res. 53 (1993) 283–290.

[83] R. Clarke, H.W. Ressom, A. Wang, J. Xuan, M.C. Liu, E.A. Gehan, Y. Wang, The properties of very high dimensional data spaces: implications for exploring gene and protein expression data, Nat. Rev. Cancer 8 (2008) 37–49.

[84] Y. Wang, D.J. Miller, R. Clarke, Approaches to working in high-dimensional data spaces: gene expression microarrays, Br. J. Cancer 98 (2008) 1023–1028.

[85] R. Clarke, T.C. Skaar, K.B. Bouker, N. Davis, Y.R. Lee, J.N. Welch, F. Leonessa, Molecular and pharmacological aspects of antiestrogen resistance, J. Steroid Biochem. Mol. Biol. 76 (2001) 71–84.

[86] R. Clarke, N. Brünner, Acquired estrogen independence and antiestrogen resistance in breast cancer: estrogen receptor-driven phenotypes? Trends Endocrinol. Metab. 7 (1996) 25–35.

[87] S. Huang, Genomics, complexity and drug discovery: insights from Boolean network models of cellular regulation, Pharmacogenomics 2 (2001) 203–222.

[88] P.E. Goss, J.N. Ingle, S. Martino, N.J. Robert, H.B. Muss, M.J. Piccart, M. Castiglione, D. Tu, L.E. Shepherd, K.I. Pritchard, R.B. Livingston, N.E. Davidson, L. Norton, E.A. Perez, J.S. Abrams, D.A. Cameron, M.J. Palmer, J.L. Pater, Randomized trial of letrozole following tamoxifen as extended adjuvant therapy in receptor-positive breast cancer: updated findings from NCIC CTG MA. 17, J. Natl. Cancer Inst. 97 (2005) 1262–1271.

[89] R.F. Hashimoto, S. Kim, I. Shmulevich, W. Zhang, M.L. Bittner, E.R. Dougherty, Growing genetic regulatory networks from seed genes, Bioinformatics 20 (2004) 1241–1247.

[90] C. Wang, J. Xuan, L. Chen, P. Zhao, Y. Wang, R. Clarke, E. Hoffman, Motif-directed component analysis for regulatory network inference, BMC Bioinform. 9 (2008), doi:10.1186/1471-2105-9-S1-S21.

[91] Y. Zhang, J. Xuan, B.G. de Los Reyes, R. Clarke, H.W. Ressom, Network motif-based identification of transcription factor-target gene relationships by integrating multi-source biological data, BMC Bioinform. 9 (2008) 203.

[92] I.B. Jeffery, S.F. Madden, P.A. McGettigan, G. Perriere, A.C. Culhane, D.G. Higgins, Integrating transcription factor binding site information with gene expression datasets, Bioinformatics 23 (2007) 298–305.
[93] L. Chen, J. Xuan, C. Wang, L.-M. Shih, Y. Wang, Z. Zhang, E.P. Hoffman, R. Clarke, Knowledge guided multi-scale independent component analysis for biomarker identification, BMC Bioinform. 9 (2008), doi:10.1186/1471-2105-9-416.
[94] S. McKenzie, N. Kyprianou, Apoptosis evasion: the role of survival pathways in prostate cancer progression and therapeutic resistance, J. Cell Biochem. 97 (2006) 18–32.
[95] Z. Gu, R.Y. Lee, T.C. Skaar, K.B. Bouker, J.N. Welch, J. Lu, A. Liu, Y. Zhu, N. Davis, F. Leonessa, N. Brunner, Y. Wang, R. Clarke, Association of interferon regulatory factor-1, nucleophosmin, nuclear factor-kappaB, and cyclic AMP response element binding with acquired resistance to faslodex (ICI 182, 780), Cancer Res. 62 (2002) 3428–3437.
[96] K.B. Bouker, T.C. Skaar, R. Riggins, D.S. Harburger, D.R. Fernandez, A. Zwart, A. Wang, R. Clarke, Interferon regulatory factor-1 (IRF-1) exhibits tumor suppressor activities in breast cancer associated with caspase activation and induction of apoptosis, Carcinogenesis 26 (2005) 1527–1535.
[97] K.B. Bouker, T.C. Skaar, D.S. Harburger, R. Riggins, D.R. Fernandez, A. Zwart, R. Clarke, The A4396G polymorphism in interferon regulatory factor-1 is frequently expressed in breast cancer, Cancer Genet. Cytogenet. 175 (2007) 61–64.
[98] D. Ron, Translational control in the endoplasmic reticulum stress response, J. Clin. Invest. 110 (2002) 1383–1388.
[99] K. Zhang, R.J. Kaufman, The unfolded protein response: a stress signaling pathway critical for health and disease, Neurology 66 (2006) S102–S109.
[100] J.B. DuRose, A.B. Tam, M. Niwa, Intrinsic capacities of molecular sensors of the unfolded protein response to sense alternate forms of endoplasmic reticulum stress, Mol. Biol. Cell 17 (2006) 3095–3107.
[101] H. Yoshida, T. Matsui, A. Yamamoto, T. Okada, K. Mori, XBP1 mRNA is induced by ATF6 and spliced by IRE1 in response to ER stress to produce a highly active transcription factor, Cell 107 (2001) 881–891.
[102] M. Ogata, S. Hino, A. Saito, K. Morikawa, S. Kondo, S. Kanemoto, T. Murakami, M. Taniguchi, I. Tanii, K. Yoshinaga, S. Shiosaka, J.A. Hammarback, F. Urano, K. Imaizumi, Autophagy is activated for cell survival after endoplasmic reticulum stress, Mol. Cell Biol. 26 (2006) 9220–9231.
[103] D.E. Feldman, V. Chauhan, A.C. Koong, The unfolded protein response: a novel component of the hypoxic stress response in tumors, Mol. Cancer Res. 3 (2005) 597–605.
[104] I.M. Clauss, M. Chu, J.L. Zhao, L.H. Glimcher, The basic domain/leucine zipper protein hXBP-1 preferentially binds to and transactivates CRE-like sequences containing an ACGT core, Nucleic Acids Res. 24 (1996) 1855–1864.
[105] H.C. Liou, M.R. Boothby, P.W. Finn, R. Davidon, N. Nabavi, L. Zeleznik, J.P. Ting, L.H. Glimcher, A new member of the leucine zipper class of proteins that binds to the HLA DR alpha promoter, Science 247 (1990) 1581–1584.
[106] R. Sriburi, S. Jackowski, K. Mori, J.W. Brewer, XBP1: a link between the unfolded protein response, lipid biosynthesis, and biogenesis of the endoplasmic reticulum, J. Cell Biol. 167 (2004) 35–41.
[107] A.H. Lee, N.N. Iwakoshi, K.C. Anderson, L.H. Glimcher, Proteasome inhibitors disrupt the unfolded protein response in myeloma cells, Proc. Natl. Acad. Sci. U.S.A. 100 (2003) 9946–9951.
[108] L. Ding, J. Yan, J. Zhu, H. Zhong, Q. Lu, Z. Wang, C. Huang, Q. Ye, Ligand-independent activation of estrogen receptor alpha by XBP-1, Nucleic Acids Res. 31 (2003) 5266–5274.
[109] A.M. Reimold, N.N. Iwakoshi, J. Manis, P. Vallabhajosyula, E. Szomolanyi-Tsuda, E.M. Gravallese, D. Friend, M.J. Grusby, F. Alt, L.H. Glimcher, Plasma cell differentiation requires the transcription factor XBP-1, Nature 412 (2001) 300–307.
[110] I.M. Clauss, E.M. Gravallese, J.M. Darling, F. Shapiro, M.J. Glimcher, L.H. Glimcher, In situ hybridization studies suggest a role for the basic region—leucine zipper protein hXBP-1 in exocrine gland and skeletal development during mouse embryogenesis, Dev. Dyn. 197 (1993) 146–156.
[111] A.M. Reimold, A. Etkin, I. Clauss, A. Perkins, D.S. Friend, J. Zhang, H.F. Horton, A. Scott, S.H. Orkin, M.C. Byrne, M.J. Grusby, L.H. Glimcher, An essential role in liver development for transcription factor XBP-1, Genes Dev. 14 (2000) 152–157.
[112] D.Y. Wang, R. Fulthorpe, S.N. Liss, E.A. Edwards, Identification of estrogen-responsive genes by complementary deoxyribonucleic acid microarray and characterization of a novel early estrogen-induced gene: EEIG1, Mol. Endocrinol. 18 (2004) 402–411.
[113] S. Tozlu, I. Girault, S. Vacher, J. Vendrell, C. Andrieu, F. Spyratos, P. Cohen, R. Lidereau, I. Bieche, Identification of novel genes that co-cluster with estrogen receptor alpha in breast tumor biopsy specimens, using a large-scale real-time reverse transcription-PCR approach, Endocr. Relat. Cancer 13 (2006) 1109–1120.
[114] C.L. Willman, C.E. Sever, M.G. Pallavicini, H. Harada, N. Tanaka, M.L. Slovak, H. Yamamoto, K. Harada, T.C. Meeker, A.F. List, T. Taniguchi, Deletion of IRF-1, mapping to chromosome 5q31.1, in human leukemia and preleukemic myelodysplasia, Science 259 (1993) 965–971.
[115] H.W. van den Berg, W.J. Leahey, M. Lynch, R. Clarke, J. Nelson, Recombinant human interferon alpha increases oestrogen receptor expression in human breast cancer cells (ZR-75-1) and sensitises them to the anti-proliferative effects of tamoxifen, Br. J. Cancer 55 (1987) 255–257.
[116] T. Matsuyama, T. Kimura, M. Kitagawa, K. Pfeffer, T. Kawakami, N. Watanabe, T.M. Kundig, R. Amakawa, K. Kishihara, A. Wakeham, J. Potter, C.L. Furlonger, A. Narendran, H. Suzuki, P.S. Ohashi, C.J. Paige, T. Taniguchi, T.W. Mak, Targeted disruption of IRF-1 or IRF-2 results in abnormal type I IFN gene induction and aberrant lymphocyte development, Cell 75 (1993) 83–97.
[117] S. Namiki, T. Nakamura, S. Oshima, M. Yamazaki, Y. Sekine, K. Tsuchiya, R. Okamoto, T. Kanai, M. Watanabe, IRF-1 mediates upregulation of LMP7 by IFN-gamma and concerted expression of immunosubunits of the proteasome, FEBS Lett. 579 (2005) 2781–2787.
[118] S.H. Lee, J.W. Kim, S.H. Oh, Y.J. Kim, S.B. Rho, K. Park, K.L. Park, J.H. Lee, IFN-gamma/IRF-1-induced p27kip1 down-regulates telomerase activity and human telomerase reverse transcriptase expression in human cervical cancer, FEBS Lett. 579 (2005) 1027–1033.
[119] S.H. Lee, J.W. Kim, H.W. Lee, Y.S. Cho, S.H. Oh, Y.J. Kim, C.H. Jung, W. Zhang, J.H. Lee, Interferon regulatory factor-1 (IRF-1) is a mediator for interferon-gamma induced attenuation of telomerase activity and human telomerase reverse transcriptase (hTERT) expression, Oncogene 22 (2003) 381–391.
[120] T. Tamura, M. Ishihara, M.S. Lamphier, N. Tanaka, I. Oishi, S. Alzawa, T. Matsuyama, T.W. Mak, S. Taki, T. Taniguchi, An IRF-1-dependent pathway of DNA damage-induced apoptosis in mitogen-activated T lymphocytes, Nature 376 (1995) 596–599.
[121] S. Prost, C.O.C. Bellamy, D.S. Cunningham, D.J. Harrison, Altered DNA repair and dysregulation of p53 in IRF-1 null hepatocytes, FASEB J. 12 (1998) 181–188.
[122] H. Nozawa, E. Oda, K. Nakao, M. Ishihara, S. Ueda, T. Yokochi, K. Ogasawara, Y. Nakatsuru, K. Hioki, S. Aizawa, T. Ishikawa, M. Katsuki, T. Muto, T. Taniguchi, N. Tanaka, Loss of transcription factor IRF-1 affects tumor susceptibility in mice carrying the Ha-ras transgene or nullizygosity for p53, Genes Dev. 1–3 (1999) 1240–1245.
[123] N. Tanaka, M. Ishihara, M. Kitagawa, H. Harada, T. Kimura, T. Matsuyama, M.S. Lamphier, S. Aizawa, T.W. Mak, T. Taniguchi, Cellular commitment to oncogene-induced transformation or apoptosis is dependent on the transcription factor IRF-1, Cell 77 (1994) 829–839.
[124] N. Tanaka, M. Ishihara, M.S. Lamphier, H. Nozawa, T. Matsuyama, T.W. Mak, S. Aizawa, T. Tokino, M. Oren, T. Taniguchi, Cooperation of the tumour suppressors IRF-1 and p53 in response to DNA damage, Nature 382 (1996) 816–818.
[125] A. Moro, A. Santos, M.J. Arana, S.E. Perea, Activation of the human p27(Kip1) promoter by IFNalpha 2b, Biochem. Biophys. Res. Commun. 269 (2000) 31–34.
[126] J. Sanceau, J. Hiscott, O. Delattre, J. Wietzerbin, IFN-beta induces serine phosphorylation of Stat-1 in Ewing's sarcoma cells and mediates apoptosis via induction of IRF-1 and activation of caspase-7, Oncogene 19 (2000) 3372–3383.
[127] K. Suk, I. Chang, Y.H. Kim, S. Kim, J.Y. Kim, H. Kim, M.S. Lee, Interferon gamma (IFNgamma) and tumor necrosis factor alpha synergism in ME-180 cervical cancer cell apoptosis and necrosis. IFNgamma inhibits cytoprotective NF-kappa B through STAT1/IRF-1 pathways, J. Biol. Chem. 276 (2001) 13153–13159.
[128] W.A. Chow, J.J. Fang, J.K. Yee, The IFN regulatory factor family participates in regulation of Fas ligand gene expression in T cells, J. Immunol. 164 (2000) 3512–3518.
[129] T.C. Skaar, K.B. Bouker, R. Clarke, Interferon regulatory factor-1 (IRF-1) in breast cancer, Proc. Am. Assoc. Cancer Res. 41 (2000) 428.
[130] K.B. Bouker, T.C. Skaar, R. Clarke, IRF-1 as a mediator of responsiveness to antiestrogens in breast cancer, Proc. Am. Assoc. Cancer Res. 41 (2000) 427–428.
[131] Z. Gu, J. Hanfelt, C. Hurley, H. Xiao, F. Gray, D. Flessate-Harley, R. Clarke, High throughput gene expression profiles associated with antiestrogen responsive *vs.* resistant breast cancer cells, Proc. Am. Assoc. Cancer Res. 40 (1999) 158–159.
[132] P.K.M. Kim, M. Armstrong, Y. Liu, P. Yan, B. Bucher, B.S. Zuckerbraun, A. Gambotto, T.R. Billiar, J.H. Yim, IRF-1 expression induces apoptosis and inhibits tumor growth in mouse mammary cancer cells in vitro and in vivo, Oncogene 23 (2004) 1125–1135.
[133] Y. Tu, D.J. Jerry, B. Pazik, S.S. Smith, Sensitivity to DNA damage is a common component of hormone-based strategies for protection of the mammary gland, Mol. Cancer Res. 3 (2005) 435–442.
[134] M.L. Bowie, E.C. Dietze, J. Delrow, G.R. Bean, M.M. Troch, R.J. Marjoram, V.L. Seewaldt, Interferon-regulatory factor-1 is critical for tamoxifen-mediated apoptosis in human mammary epithelial cells, Oncogene 23 (2004) 8743–8755.
[135] J.H. Yim, S.H. Ro, J.K. Lowney, S.J. Wu, J. Connett, G.M. Doherty, The role of interferon regulatory factor-1 and interferon regulatory factor-2 in IFN-gamma growth inhibition of human breast carcinoma cell lines, J. Interferon Cytokine Res. 23 (2003) 501–511.
[136] E. Pizzoferrato, Y. Liu, A. Gambotto, M.J. Armstrong, M.T. Stang, W.E. Gooding, S.M. Alber, S.H. Shand, S.C. Watkins, W.J. Storkus, J.H. Yim, Ectopic expression of interferon regulatory factor-1 promotes human breast cancer cell death and results in reduced expression of survivin, Cancer Res. 64 (2004) 8381–8388.
[137] W.F. Tam, R. Sen, IκB family members function by different mechanisms, J. Biol. Chem. 276 (2001) 7701–7704.
[138] A. Yaron, A. Hatzubai, M. Davis, I. Lavon, S. Amit, A.M. Manning, J.S. Andersen, M. Mann, F. Mercurio, Y. Ben Neriah, Identification of the receptor component of the IkappaBalpha-ubiquitin ligase, Nature 396 (1998) 590–594.
[139] V. Bours, M. Bentires-Alj, A.C. Hellin, P. Viatour, P. Robe, S. Delhalle, V. Benoit, M.P. Merville, Nuclear factor-kappa B, cancer, and apoptosis, Biochem. Pharmacol. 60 (2000) 1085–1089.
[140] H. Nakshatri, P. Bhat-Nakshatri, D.A. Martin, R.J. Goulet, G.W. Sledge, Constitutive activation of NF-kappaB during progression of breast cancer to hormone-independent growth, Mol. Cell Biol. 17 (1997) 3629–3639.
[141] D.K. Biswas, A.P. Cruz, E. Gansberger, A.B. Pardee, Epidermal growth factor-induced nuclear factor kappa B activation: a major pathway of cell-cycle

progression in estrogen-receptor negative breast cancer cells, Proc. Natl. Acad. Sci. U.S.A. 97 (2000) 8542–8547.
[142] R.B. Dickson, M.E. Lippman, Growth factors in breast cancer, Endocr. Rev. 16 (1995) 559–589.
[143] R.W. Clarkson, C.J. Watson, NF-kappaB and apoptosis in mammary epithelial cells, J. Mammary Gland Biol. Neoplasia 4 (1999) 165–175.
[144] D.W. Kim, M.A. Sovak, G. Zanieski, G. Nonet, R. Romieu-Mourez, A.W. Lau, L.J. Hafer, P. Yaswen, M. Stampfer, A.E. Rogers, J. Russo, G.E. Sonenshein, Activation of NF-kappaB/Rel occurs early during neoplastic transformation of mammary cells, Carcinogenesis 21 (2000) 871–879.
[145] S. Tonko-Geymayer, W. Doppler, An essential link to mammary cancer? Nat. Med. 8 (2002) 108–110.
[146] P.C. Cogswell, D.C. Guttridge, W.K. Funkhouser, A.S. Baldwin Jr., Selective activation of NF-kappa B subunits in human breast cancer: potential roles for NF-kappa B2/p52 and for Bcl-3, Oncogene 19 (2000) 1123–1131.
[147] M.A.C. Pratt, T.E. Bishop, D. White, G. Yasvinski, M. Menard, M.Y. Niu, R. Clarke, Estrogen withdrawal-induced NF-kappaB activity and bcl-3 expression in breast cancer cells: roles in growth and hormone independence, Mol. Cell Biol. 23 (2003) 6887–6900.
[148] N. Brünner, V. Boulay, A. Fojo, C. Freter, M.E. Lippman, R. Clarke, Acquisition of hormone-independent growth in MCF-7 cells is accompanied by increased expression of estrogen-regulated genes but without detectable DNA amplifications, Cancer Res. 53 (1993) 283–290.
[149] S.P. Hehner, T.G. Hofmann, W. Droge, M.L. Schmitz, The antiinflammatory sesquiterpene lactone parthenolide inhibits NF- kappa B by targeting the I kappa B kinase complex, J. Immunol. 163 (1999) 5617–5623.
[150] Y. Zhu, A. Wang, M.C. Liu, A. Zwart, R.Y. Lee, A. Gallagher, Y. Wang, W.R. Miller, J.M. Dixon, R. Clarke, Estrogen receptor alpha (ER) positive breast tumors and breast cancer cell lines share similarities in their transcriptome data structures, Int. J. Oncol. 29 (2006) 1581–1589.
[151] C.J. Creighton, K.E. Cordero, J.M. Larios, R.S. Miller, M.D. Johnson, A.M. Chinnaiyan, M.E. Lippman, J.M. Rae, Genes regulated by estrogen in breast tumor cells in vitro are similarly regulated in vivo in tumor xenografts and human breast tumors, Genome Biol. 7 (2006) R28.
[152] Y. Zhu, B. Singh, S. Hewitt, A. Liu, B. Gomez, A. Wang, R. Clarke, Expression patterns among interferon regulatory factor-1, human X-box binding protein-1, nuclear factor kappa B, nucleophosmin, estrogen receptor alpha and progesterone receptor proteins in breast cancer tissue microarrays, Int. J. Oncol. 28 (2006) 67–76.
[153] A. de la Fuente, N. Bing, I. Hoeschele, P. Mendes, Discovery of meaningful associations in genomic data using partial correlation coefficients, Bioinformatics 20 (2004) 3565–3574.
[154] M.P. Davies, D.L. Barraclough, C. Stewart, K.A. Joyce, R.M. Eccles, R. Barraclough, P.S. Rudland, D.R. Sibson, Expression and splicing of the unfolded protein response gene XBP-1 are significantly associated with clinical outcome of endocrine-treated breast cancer, Int. J. Cancer 123 (2008) 85–88.
[155] T.C. Skaar, S.C. Prasad, S. Sharaeh, M.E. Lippman, N. Brünner, R. Clarke, Two-dimensional gel electrophoresis analyses identify nucleophosmin as an estrogen regulated protein associated with acquired estrogen-independence in human breast cancer cells, J. Steroid Biochem. Mol. Biol. 67 (1998) 391–402.
[156] T. Kondo, N. Minamino, T. Nagamura-Inoue, M. Matsumoto, T. Taniguchi, N. Tanaka, Identification and characterization of nucleophosmin/B23/numatrin which binds the anti-oncogenic transcription factor IRF-1 and manifests oncogenic activity, Oncogene 15 (1997) 1275–1281.
[157] B. Brankin, T.C. Skaar, B.J. Trock, M. Berris, R. Clarke, Autoantibodies to numatrin: an early predictor for relapse in breast cancer, Cancer Epidemiol. Biomarkers Prev. 7 (1998) 1109–1115.
[158] M. Saura, C. Zaragoza, C. Bao, A. McMillan, C.J. Lowenstein, Interaction of interferon regulatory factor-1 and nuclear factor kappaB during activation of inducible nitric oxide synthase transcription, J. Mol. Biol. 289 (1999) 459–471.
[159] P.D. Drew, G. Franzoso, K.G. Becker, V. Bours, L.M. Carlson, U. Siebenlist, K. Ozato, NF kappa B and interferon regulatory factor 1 physically interact and synergistically induce major histocompatibility class I gene expression, J. Interferon Cytokine Res. 15 (1995) 1037–1045.
[160] M. Karin, Y. Cao, F.R. Greten, Z.W. Li, NF-kappaB in cancer: from innocent bystander to major culprit, Nat. Rev. Cancer 2 (2002) 301–310.

Breast Cancer Res Treat
DOI 10.1007/s10549-009-0509-8

BRIEF REPORT

Frequent loss of heterozygosity at the interferon regulatory factor-1 gene locus in breast cancer

Luciane R. Cavalli · Rebecca B. Riggins · Antai Wang · Robert Clarke · Bassem R. Haddad

Received: 6 August 2009 / Accepted: 8 August 2009

Abstract The interferon regulatory factor-1 (*IRF1*) gene, localized on chromosome 5q31.1, is mutated or rearranged in several cancers including some hematopoietic and gastric cancers. However, whether loss of *IRF1* occurs in sporadic breast cancer is unknown. Loss of 5q12-31 is reported in 11% of sporadic breast cancers, and high-resolution array-CGH studies have shown loss at 5q31.1 in 50% of breast cancers with a mutated *BRCA1* gene. Functionally, overexpression of *IRF1* reduces, and a dominant negative *IRF1* construct increases, tumorigenesis of human breast cancer xenografts. Taken together, these observations indicate that the *IRF1* gene may play a potentially important role as a breast cancer tumor suppressor gene. In this study, we investigated allelic loss of the *IRF1* gene in breast tumor specimens from 52 women with invasive breast cancer using an *IRF1* intragenic dinucleotide polymorphic marker. Thirty-seven cases were informative. LOH at the *IRF1* locus was detected in 32% of these informative cases (12/37). There was a significant association between *IRF1* loss and both older age ($P = 0.0167$) and earlier stage (Stages 1 and 2) ($P = 0.0165$). To assess the association of IRF1 mRNA expression with clinical outcomes in breast cancer, we studied data from two published gene expression microarray datasets. In breast cancer patients, low IRF1 mRNA expression is strongly correlated with both risk of recurrence (OR = 3.00; $P = 0.003$; $n = 273$ cases) and risk of death (OR = 4.18; $P = 0.004$; $n = 191$ cases). Our findings strongly imply a tumor suppressor role for the *IRF1* gene in breast cancer.

Keywords Interferon regulatory factor-1 · *IRF1* · Breast cancer · Disease survival · Tumor suppressor gene

Introduction

The interferon regulatory factor-1 (*IRF1*) gene mediates interferon and other cytokine effects and exhibits antitumor activity in vivo and in vitro. *IRF1* can also reverse the oncogenic transformation of cells induced by the overexpression of both *RAS* and *MYC* in mouse models [1]. Since functional roles for *RAS* and *MYC* are established in human breast cancer [2, 3], a loss of *IRF1* function might be important in this disease. Functionally, overexpression of *IRF1* reduces [4, 5], and a dominant negative *IRF1* construct increases [4], tumorigenesis of human breast cancer xenografts. We and others have identified *IRF1* as a key regulator of breast cancer cell survival [5–8] that can activate a caspase cascade [4] and induce apoptosis [5, 6]. More specifically, the proapoptotic effects of *IRF1* include activation/regulation of caspases-1 [9], -3 [4], -7 [4, 10], and -8 [4, 11]. *IRF1* can also induce apoptosis through both

L. R. Cavalli · R. B. Riggins · R. Clarke · B. R. Haddad
Department of Oncology, Lombardi Comprehensive Cancer Center, Georgetown University School of Medicine, Washington, DC 20057, USA

A. Wang
Department of Biostatistics and Bioinformatics, Lombardi Comprehensive Cancer Center, Georgetown University School of Medicine, Washington, DC 20057, USA

B. R. Haddad
Department of Obstetrics and Gynecology, Lombardi Comprehensive Cancer Center, Georgetown University School of Medicine, Washington, DC 20057, USA

R. Clarke (✉)
Departments of Oncology, Georgetown University School of Medicine, Room W405A Research Building, 3970 Reservoir Rd, NW, Washington, DC 20007, USA
e-mail: clarker@georgetown.edu

Journal : Large 10549 | Dispatch : 13-8-2009 | Pages : 5
Article No. : 509 | LE | TYPESET
MS Code : BREA4520 | CP | DISK

TP53-dependent and *TP53*-independent signaling [9, 12]. *TP53* is often mutated in breast cancer [13], and many breast tumors initially respond to drugs and hormones through *TP53*-dependent and *TP53*-independent signaling. We have shown that *IRF1* is a key determinant of responsiveness to antiestrogen therapies in breast cancer [6, 7].

Whether *IRF1* is a true tumor suppressor gene (TSG) in breast cancer is unknown. Established TSGs often show evidence of homozygous or heterozygous gene loss. For instance, while loss of *BRCA1* function in inherited breast cancers is usually a consequence of gene mutation(s), loss of *BRCA1* expression in sporadic breast cancers is often the result of loss of heterozygosity (LOH) accompanied by hypermethylation of a CpG island in the 5′ region close to the transcription start site of the remaining allele [14, 15]. *IRF1* has been implicated as a putative TSG in leukemias and preleukemic myelodysplasias, and *IRF1* is either mutated or rearranged in some hematopoietic disorders [16]. *IRF1* was shown to be the true target of frequent deletions (LOH) in esophageal cancer [17] and gastric cancer [18]. *IRF1* is located at 5q31.1, a region shown to be commonly lost in two large studies evaluating breast tumors by chromosomal comparative genomic hybridization (CGH). Deletion of 5q12-31 is detected in 11% of sporadic breast cancers [19] and 5q deletion is seen in 86% of *BRCA1* tumors [20]. More recently, a high-resolution array-CGH study has shown loss at 5q31.1 in 50% of *BRCA1*-positive breast cancers [21]. Whether loss at the *IRF1* locus is the driver in these cancers and whether *IRF1* gene loss occurs in sporadic breast cancers are unknown.

IRF1 is inactivated by a point mutation in gastric cancer, suggesting that the loss of function of *IRF1* may be critical for the development of this disease [18]. When attempting to generate an *IRF1* riboprobe from MCF-7 breast cancer cells mRNA, we found a single nucleotide polymorphism (SNP) in the *IRF1* coding region [22] and a novel *IRF1* splice variant (K. B. Bouker; unpublished observations). The *IRF1* A4396G SNP is more frequent in human breast cancer cell lines than in the general population and is more frequently expressed in African-American than Caucasian subjects [22]. It is not known whether *IRF1* A4396G contributes to the earlier age [23] or higher stage at diagnosis [24] or the lower overall survival rate of African-American compared with non-Hispanic white and Hispanic women [25]. When considered together, these observations strongly suggest that *IRF1* may be a TSG in breast cancer. Since one of the key features of TSGs is somatic loss, we designed this study to determine the incidence of *IRF1* loss in a series of 52 invasive breast tumors. Considering that *IRF1* LOH might be expected to reduce mRNA expression, we also explored whether low *IRF1* mRNA expression is associated with poor clinical outcomes in breast cancer patients.

Methods

We determined *IRF1* loss by LOH in breast tissue specimens from 52 patients with sporadic breast cancer obtained from the tumor bank at the Lombardi Comprehensive Cancer Center (LCCC). In each case, a paraffin block with breast tumor tissues and a second block with normal tissues (skin, negative lymph node, or a normal breast tissue block) were identified. An H&E-stained slide from each block was evaluated by a breast pathologist to confirm the diagnosis and mark the areas with malignant tissue or normal tissue. A 100-μm consecutive section was obtained from each block, and the tissues of interest were grossly microdissected with a razor blade to isolate malignant cells. Corresponding normal cells from a different block were obtained for each case. DNA was extracted from the tissue using the DNeasy QIAGEN kit according to manufacturer's instructions (QIAGEN Inc. Valencia, California).

To study LOH at the *IRF1* locus, we selected an intragenic, dinucleotide, polymorphic marker (*IRF1* Dinucleotide Repeat, Allele Set GDB: 211036), with a high degree of heterozygosity (74% heterozygosity). The sequences of the oligonucleotide primers were obtained from the Genome Data Base (GDB) (http://www.gdb.org): Forward: 5′-ATGGCAGATAGGTCCACCGG-3′/Reverse: 5′-TCAT CCTCATCTGTTGTACG-3′. Primers were fluorescently labeled and PCR amplification was performed using a standard protocol. Allele sizes were determined by electrophoresis of PCR products in 6% denaturing polyacrylamide gels and were compared to ROX 500 size standards (Applied Biosystems, Foster City, CA), using an automated sequencer (ABI 377), according to the manufacturer's instructions (Applied Biosystems). Fluorescent signals from the different size alleles were recorded and analyzed using GENOTYPER version 2.1 and GENESCAN version 3.1 software (Applied Biosystems). Following visual examination of computer printouts by two independent observers, LOH was determined mathematically according to the Genotyper User's Manual (Applied Biosystems).

The publicly available ONCOMINE cancer gene expression microarray database [www.oncomine.org; 26] was used to search for relationships between *IRF1* mRNA expression and outcomes in breast cancer clinical studies. Normalized Affymetrix MAS 5.0 gene expression data, originally published by Wang et al. [27] and Desmedt et al. [28], were downloaded from ONCOMINE. Statistical analysis was done using SigmaStat (Systat Software, Inc., San Jose, CA) and S-PLUS (Insightful, Seattle, WA). Median *IRF1* expression values across all samples, and between the top and bottom quartiles of expression, were compared by Mann–Whitney Rank-Sum test, and odds ratios were calculated for the association of low *IRF1* expression with poor outcomes following logistic

Journal : Large 10549 | Dispatch : 13-8-2009 | Pages : 5
Article No. : 509 | ☐ LE | ☐ TYPESET
MS Code : BREA4520 | ☑ CP | ☑ DISK

regression analysis. Statistical significance was defined as ≥95% confidence interval, or $\alpha = 0.05$.

Results

In this study, 37 cases (71%) were informative for the *IRF1* dinucleotide marker used for LOH analysis. LOH was detected in 12 of these informative cases (12/37; 32%). Figure 1 shows a representative case with no LOH (Fig. 1a, bottom panel) and another representative case with LOH (Fig. 1b, bottom panel). A significant correlation was found between LOH at the *IRF1* locus and both older age (P value = 0.0167 based on a two sample t-test) and earlier stage (Stages 1 and 2) (P value = 0.0165 based on Fisher's exact test).

The data from two published gene expression microarray datasets were assessed to investigate mRNA expression and clinical outcome [27, 28]. In both studies, median *IRF1* mRNA expression, as detected by Affymetrix U133A GeneChips, was significantly reduced in patients with a worse outcome (recurrence, $P = 0.004$; death, $P = 0.021$). Moreover, when these expression values data were separated into quartiles and analyzed by logistic regression (Table 1), low *IRF1* mRNA expression (first compared with fourth quartile) was significantly associated with both recurrence (OR = 3.00, $P = 0.003$) and death (OR = 4.18, $P = 0.004$).

Discussion

In our study, a significant correlation was found between LOH at the *IRF1* locus and both older age ($P = 0.0167$) and earlier stage (Stages 1 and 2) ($P = 0.0165$). An inverse association between IRF1 protein expression and tumor grade has been reported [29], and IRF1 protein levels are lower in breast tumors than in adjacent normal cells [30]. However, subcellular location also affects IRF1 correlation with clinical measures, and these correlations would not be apparent with LOH measurements. For example, cytosolic IRF-1 protein (but not nuclear *IRF1*) is associated with age (similar to LOH findings) and ER expression, whereas nuclear IRF1 protein expression (but not cytosolic *IRF1*) correlates with PgR expression [31].

Since *IRF1* LOH might be expected to reduce mRNA expression, we explored the association of *IRF1* mRNA with the key measures of recurrence status (recurrent; non-recurrent) and vital status (alive; dead). Using the publicly available ONCOMINE cancer gene expression microarray database [26], we searched for relationships between *IRF1* mRNA expression and outcomes from two breast cancer clinical studies. The Wang et al. study includes 273 women diagnosed with lymph node-negative breast cancer who had not received systemic adjuvant therapy, but whose tumors are representative of a wide range of clinical/pathological features (including age, stage, and tumor size). The original goal of that study was to identify a gene expression signature that could predict recurrence of metastatic breast cancer in women with node-negative disease [27]. The more recent Desmedt et al. study is an independent validation of Wang et al. and includes 191 untreated patients less than 61 years of age with node-negative, T1 and T2 breast cancer [28]. The primary endpoint for the study was overall survival, and median follow-up time was 13.6 years. While *IRF1* gene copy number data and protein expression data were not available from these studies, we could assess the potential of *IRF1* to act as a TSG as predicted by the likelihood that low *IRF1* mRNA expression was associated with poor clinical outcomes. We have observed that low IRF1 mRNA expression is strongly correlated with both risk of recurrence and risk of death. Studies to determine directly the role of *IRF1* LOH in these outcomes are in progress.

IRF1 LOH in breast cancer may reflect haploinsufficiency; over 35 TSGs are known in which haploinsufficiency accounts for their contribution to carcinogenesis (though not necessarily breast carcinogenesis) [32, 33]. Reduced IRF1 activity in experimental breast cancer models is functionally associated with increased cell survival,

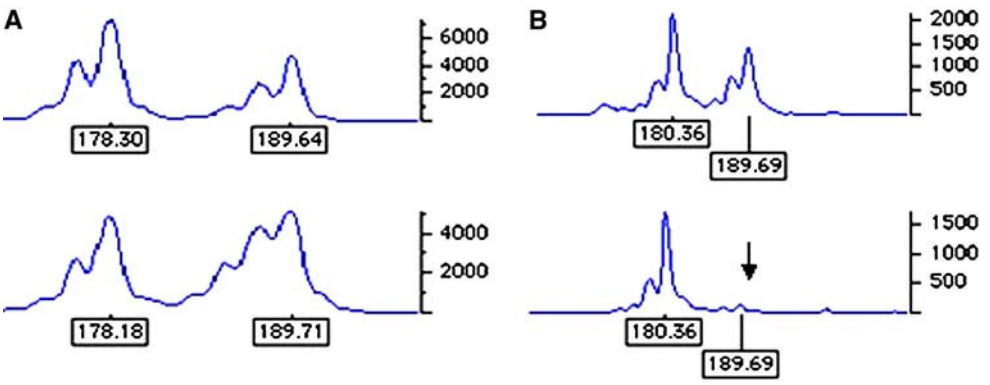


Fig. 1 LOH analysis at the *IRF1* locus of two representative breast cancer cases. **a** A case with no LOH. **b** A case with LOH. (In **a**, **b**, *Top panels* shows the analysis performed in normal cells and *bottom panels* the analysis performed in tumor cells.)

Journal : Large 10549 | Dispatch : 13-8-2009 | Pages : 5
Article No. : 509 | ☐ LE | ☐ TYPESET
MS Code : BREA4520 | ☑ CP | ☑ DISK

Table 1 *IRF1* mRNA expression and clinical outcomes

Study	Endpoint	*IRF1* (median)	*IRF1* (quartile)	Odds ratio[a]
Wang et al.	No disease ($n = 180$)	$P = 0.004$	First = 0.971	OR = 3.00
				CI = 1.44–6.27
	Recurrence ($n = 93$)		Fourth = 0.455	$P = 0.003$
Desmedt et al.	Alive ($n = 135$)	$P = 0.021$	First = 0.905	OR = 4.18
				CI = 1.56–11.21
	Dead ($n = 56$)		Fourth = 0.355	$P = 0.004$

[a] Odds ratios were calculated for the first (highest) versus fourth (lowest) quartiles of expression data, and denote the association of low *IRF1* mRNA expression with poor outcome (recurrence or death); CI = 95% confidence interval

reduced caspase activation, and apoptosis (4–6). While LOH may not be the only contributor to low *IRF1* expression [34, 35], the strong association of low *IRF1* mRNA and recurrence status and survival imply an important TSG role for *IRF1* in many sporadic breast cancers.

Acknowledgments This study was supported in part by a pilot grant from the Jess and Mildred Fisher Center for Familial Cancer Research at Georgetown University. We thank the Histopathology and Tissue Shared Resource (HTSR) of the Lombardi Comprehensive Cancer Center, which is partially supported by National Institutes of Health Grant 1P30-CA-51008 (Cancer Center Support Grant, CCSG), for providing the histopathology support.

References

1. Nozawa H, Oda E, Nakao K, Ishihara M, Ueda S, Yokochi T, Ogasawara K, Nakatsuru Y, Hioki K, Aizawa S, Ishikawa T, Katsuki M et al (1999) Loss of transcription factor IRF-1 affects tumor susceptibility in mice carrying the Ha-ras transgene or nullizygosity for p53. Genes Dev 1–3:1240–1245
2. Bos JL (1989) Ras oncogenes in human cancer: a review. Cancer Res 49:4682–4689
3. Deming SL, Nass SJ, Dickson RB, Trock BJ (2000) C-myc amplification in breast cancer: a meta-analysis of its occurrence and prognostic relevance. Br J Cancer 83:1688–1695
4. Bouker KB, Skaar TC, Riggins R, Harburger DS, Fernandez DR, Zwart A, Wang A, Clarke R (2005) Interferon regulatory factor-1 (IRF-1) exhibits tumor suppressor activities in breast cancer associated with caspase activation and induction of apoptosis. Carcinogenesis 26:1527–1535
5. Kim PKM, Armstrong M, Liu Y, Yan P, Bucher B, Zuckerbraun BS, Gambotto A, Billiar TR, Yim JH (2004) IRF-1 expression induces apoptosis and inhibits tumor growth in mouse mammary cancer cells in vitro and in vivo. Oncogene 23:1125–1135
6. Gu Z, Lee RY, Skaar TC, Bouker KB, Welch JN, Lu J, Liu A, Zhu Y, Davis N, Leonessa F, Brunner N, Wang Y et al (2002) Association of interferon regulatory factor-1, nucleophosmin, nuclear factor-kappaB, and cyclic AMP response element binding with acquired resistance to faslodex (ICI 182, 780). Cancer Res 62:3428–3437
7. Bouker KB, Skaar TC, Fernandez DR, O'Brien KA, Clarke R (2004) Interferon regulatory factor-1 mediates the proapoptotic but not cell cycle arrest effects of the steroidal antiestrogen ICI 182, 780 (Faslodex, Fulvestrant). Cancer Res 64:4030–4039
8. Bowie ML, Dietze EC, Delrow J, Bean GR, Troch MM, Marjoram RJ, Seewaldt VL (2004) Interferon-regulatory factor-1 is critical for tamoxifen-mediated apoptosis in human mammary epithelial cells. Oncogene 23:8743–8755
9. Tamura T, Ishihara M, Lamphier MS, Tanaka N, Oishi I, Alzawa S, Matsuyama T, Mak TW, Taki S, Taniguchi T (1995) An IRF-1-dependent pathway of DNA damage-induced apoptosis in mitogen-activated T lymphocytes. Nature 376:596–599
10. Sanceau J, Hiscott J, Delattre O, Wietzerbin J (2000) IFN-beta induces serine phosphorylation of stat-1 in Ewing's sarcoma cells and mediates apoptosis via induction of IRF-1 and activation of caspase-7. Oncogene 19:3372–3383
11. Suk K, Chang I, Kim YH, Kim S, Kim JY, Kim H, Lee MS (2001) Interferon gamma (IFNgamma) and tumor necrosis factor alpha synergism in ME-180 cervical cancer cell apoptosis and necrosis. IFNgamma inhibits cytoprotective NF-kappa B through STAT1/IRF-1 pathways. J Biol Chem 276:13153–13159
12. Tanaka N, Ishihara M, Lamphier MS, Nozawa H, Matsuyama T, Mak TW, Aizawa S, Tokino T, Oren M, Taniguchi T (1996) Cooperation of the tumour suppressors IRF-1 and p53 in response to DNA damage. Nature 382:816–818
13. Elledge RM, Allred DC (1994) The p53 tumor suppressor gene in breast cancer. Breast Cancer Res Treat 32:39–47
14. Esteller M, Silva JM, Dominguez G, Bonilla F, Matias-Guiu X, Lerma E, Bussaglia E, Prat J, Harkes IC, Repasky EA, Gabrielson E, Schutte M et al (2000) Promoter hypermethylation and BRCA1 inactivation in sporadic breast and ovarian tumors. J Natl Cancer Inst 92:564–569
15. Wei M, Grushko TA, Dignam J, Hagos F, Nanda R, Sveen L, Xu J, Fackenthal J, Tretiakova M, Das S, Olopade OI (2005) BRCA1 promoter methylation in sporadic breast cancer is associated with reduced BRCA1 copy number and chromosome 17 aneusomy. Cancer Res 65:10692–10699
16. Willman CL, Sever CE, Pallavicini MG, Harada H, Tanaka N, Slovak ML, Yamamoto H, Harada K, Meeker TC, List AF, Taniguchi T (1993) Deletion of IRF-1, mapping to chromosome 5q31.1, in human leukemia and preleukemic myelodysplasia. Science 259:968–971
17. Ogasawara S, Tamura G, Maesawa C, Suzuki Y, Ishida K, Satoh N, Uesugi N, Saito K, Satodate R (1996) Common deleted region on the long arm of chromosome 5 in esophageal carcinoma. Gastroenterology 110:52–57
18. Nozawa H, Oda E, Ueda S, Tamura G, Maesawa C, Muto T, Taniguchi T, Tanaka N (1998) Functionally inactivating point mutation in the tumor-suppressor IRF-1 gene identified in human gastric cancer. Int J Cancer 77:522–527
19. Tirkkonen M, Tanner M, Karhu R, Kallioniemi A, Isola J, Kallioniemi OP (1998) Molecular cytogenetics of primary breast cancer by CGH. Genes Chromosomes Cancer 21:177–184
20. Tirkkonen M, Johannsson O, Agnarsson BA, Olsson H, Ingvarsson S, Karhu R, Tanner M, Isola J, Barkardottir RB, Borg A, Kallioniemi OP (1997) Distinct somatic genetic changes

Journal : Large 10549 | Dispatch : 13-8-2009 | Pages : 5
Article No. : 509 | ☐ LE | ☐ TYPESET
MS Code : BREA4520 | ☑ CP | ☑ DISK

associated with tumor progression in carriers of BRCA1 and BRCA2 germ-line mutations. Cancer Res 57:1222–1227
21. Johannsdottir HK, Jonsson G, Johannesdottir G, Agnarsson BA, Eerola H, Arason A, Heikkila P, Egilsson V, Olsson H, Johannsson OT, Nevanlinna H, Borg A et al (2006) Chromosome 5 imbalance mapping in breast tumors from BRCA1 and BRCA2 mutation carriers and sporadic breast tumors. Int J Cancer 119:1052–1060
22. Bouker KB, Skaar TC, Harburger DS, Riggins R, Fernandez DR, Zwart A, Clarke R (2007) The A4396G polymorphism in interferon regulatory factor-1 is frequently expressed in breast cancer. Cancer Genet Cytogenet 175:61–64
23. Aziz H, Hussain F, Sohn C, Mediavillo R, Saitta A, Hussain A, Brandys M, Homel P, Rotman M (1999) Early onset of breast carcinoma in African American women with poor prognostic factors. Am J Clin Oncol 22:436–440
24. Weir HK, Thun MJ, Hankey BF, Ries LA, Howe HL, Wingo PA, Jemal A, Ward E, Anderson RN, Edwards BK (2003) Annual report to the nation on the status of cancer, 1975–2000, featuring the uses of surveillance data for cancer prevention and control. J Natl Cancer Inst 95:1276–1299
25. Shavers VL, Harlan LC, Stevens JL (2003) Racial/ethnic variation in clinical presentation, treatment, and survival among breast cancer patients under age 35. Cancer 97:134–147
26. Rhodes DR, Yu J, Shanker K, Deshpande N, Varambally R, Ghosh D, Barrette T, Pandey A, Chinnaiyan AM (2004) ONCOMINE: a cancer microarray database and integrated data-mining platform. Neoplasia 6:1–6
27. Wang Y, Klijn JG, Zhang Y, Sieuwerts AM, Look MP, Yang F, Talantov D, Timmermans M, Meijer-van Gelder ME, Yu J, Jatkoe T, Berns EM et al (2005) Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. Lancet 365:671–679
28. Desmedt C, Piette F, Loi S, Wang Y, Lallemand F, Haibe-Kains B, Viale G, Delorenzi M, Zhang Y, d'Assignies MS, Bergh J, Lidereau R et al (2007) Strong time dependence of the 76-gene prognostic signature for node-negative breast cancer patients in the TRANSBIG multicenter independent validation series. Clin Cancer Res 13:3207–3214
29. Connett JM, Badri L, Giordano TJ, Connett WC, Doherty GM (2005) Interferon regulatory factor 1 (IRF-1) and IRF-2 expression in breast cancer tissue microarrays. J Interferon Cytokine Res 25:587–594
30. Doherty GM, Boucher L, Sorenson K, Lowney J (2001) Interferon regulatory factor expression in human breast cancer. Ann Surg 233:623–629
31. Zhu Y, Singh B, Hewitt S, Liu A, Gomez B, Wang A, Clarke R (2006) Expression patterns among interferon regulatory factor-1, human X-box binding protein-1, nuclear factor kappa B, nucleophosmin, estrogen receptor alpha and progesterone receptor proteins in breast cancer tissue microarrays. Int J Oncol 28:67–76
32. Payne SR, Kemp CJ (2005) Tumor suppressor genetics. Carcinogenesis 26:2031–2045
33. Santarosa M, Ashworth A (2004) Haploinsufficiency for tumour suppressor genes: when you don't need to go all the way. Biochim Biophys Acta 1654:105–122
34. Book MM, Yu-Lee LY (2001) Sp1 is required for prolactin activation of the interferon regulatory factor-1 gene. Mol Cell Endocrinol 184:135–141
35. Li X, Leung S, Qureshi S, Darnell JE Jr, Stark GR (1996) Formation of STAT1-STAT2 heterodimers and their role in the activation of IRF-1 gene transcription by interferon-alpha. J Biol Chem 271:5790–5794

Journal : **Large 10549** | Dispatch : **13-8-2009** | Pages : **5**
Article No. : **509** | ☐ LE | ☐ TYPESET
MS Code : **BREA4520** | ☑ CP | ☑ DISK